理學博士學位論文

Structural and Functional Studies of Enhanced Intracellular Survival Proteins from *Mycobacterium tuberculosis* and *Mycobacterium smegmatis*

2014年 2月

서울 大學校 大學院

化學部 生化學專攻

김 경 훈

Structural and Functional Studies of Enhanced Intracellular Survival Proteins from *Mycobacterium tuberculosis* and *Mycobacterium smegmatis*

指導教授 徐 世 源

이 論文을 理學博士學位 論文으로 提出함

2013年 11月

서울 大學校 大學院

化學部 生化學專攻

김 경 훈

김경훈의 理學博士學位論文을 認准함

2013年 11月

委 員 長 박 승 범

副委員長 서 세 원

委 員 이 재 영

委 員 양 진 국

委 員 박 홍 도

Structural and Functional Studies of Enhanced Intracellular Survival Proteins from *Mycobacterium tuberculosis* and *Mycobacterium smegmatis*

Thesis by

Kyoung Hoon Kim

Professor: Se Won Suh

A Thesis Submitted to the Graduate Faculty of Seoul National University in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy

2014

Abstract

The intracellular pathogen *Mycobacterium tuberculosis* causes tuberculosis. Enhanced intracellular survival (Eis) protein, secreted by *M. tuberculosis*, enhances survival of *Mycobacterium smegmatis* in macrophages. *M. tuberculosis* Eis was shown to suppress host immune defenses by negatively modulating autophagy, inflammation, and cell death through c-Jun N-terminal kinase (JNK)-dependent inhibition of reactive oxygen species generation. *M. tuberculosis* Eis was demonstrated to contribute to drug resistance by acetylating multiple amines of aminoglycosides. However, the mechanism of enhanced intracellular survival by *M. tuberculosis* Eis remains unanswered. Therefore, I have characterized both *M. tuberculosis* and *M. smegmatis* Eis proteins biochemically and structurally. I have discovered that *M. tuberculosis* Eis is an efficient N^{ε}-acetyltransferase, rapidly acetylating Lys55 of dual-specificity protein phosphatase 16 (DUSP16)/mitogen-activated protein kinase phosphatase-7 (MKP-7), a JNK-specific phosphatase. In contrast, *M. smegmatis* Eis is more efficient as an N^{α}-acetyltransferase. I also show that *M. smegmatis* Eis acetylates aminoglycosides as readily as *M. tuberculosis*

Eis. Furthermore, *M. tuberculosis* Eis, but not *M. smegmatis* Eis, inhibits lipopolysaccharide-induced JNK phosphorylation. This functional difference against DUSP16/MKP-7 can be understood by comparing the structures of two Eis proteins. The active site of *M. tuberculosis* Eis with a narrow channel appears to be more suitable for sequence-specific recognition of the protein substrate than the pocket-shaped active site of *M. smegmatis* Eis. I propose that *M. tuberculosis* Eis initiates the inhibition of JNK-dependent autophagy, phagosome maturation, and reactive oxygen species generation by acetylating Lys55 of DUSP16/MKP-7. My work thus provides new insights into the mechanism of suppressing host immune responses and enhancing mycobacterial survival within macrophages by *M. tuberculosis* Eis.

Keywords: *Mycobacterium tuberculosis* / *Mycobacterium smegmatis* / Enhanced intracellular survival / MKP-7 / DUSP16 / Rv2416c

Table of Contents

3.3 Ligand binding at the active site · 39

3.4 Comparison of aminoglycosides acetyltransferase activity of Eis proteins from *M. tuberculosis* and *M. smegmatis* · · · · · · · · · · · · · · · · 42

3.5 Comparison of overall structure of Eis proteins from *M. tuberculosis* and *M. smegmatis* · 45

3.6 Identification of DUSP16/MKP-7 as the N^{ϵ}-acetylation target of *M. tuberculosis* Eis. · · · · · · · · · · · · · · · · · · ·

List of Tables

Table 1. Data collection, phasing, and refinement statistics · · · · · · · · · · · 21

Table 2. Kinetic parameters of aminoglycoside N-acetylation by Eis proteins · · · · · · · · · · ·

List of Figures

Figure 1. Survival of *M. smegmatis* containing pOLYG (control, shuttle vector) or p69 (pOLYG +*M. tuberculosis* Eis open reading frame) in U-937 cells · · ·

Abbreviations

AIDS	: acquired immunodeficiency syndrome
BMDM	: bone marrow-derived macrophage
CoA	: coenzyme A
DTT	: dithiothreitol
DUSP16	: dual-specificity protein phosphatase 16
EDTA	: ethylenediaminetetraacetic acid
Eis	: enhanced intracellular survival
ERK1/2	: extracellular signal-regulated kinase 1/2
GNAT	: GCN5-related *N*-acetyltransferase
IL-4	: interleukin-4
IL-10	: interleukin-10
JAK	: janus kinase
JNK	: c-Jun N-terminal kinase
IPTG	: isopropyl-1-thio-β-galatopyranoside
MALDI	: matrix-assisted laser desorption/ionization
MAPK	: mitogen-activated protein kinase
MKP-7	: mitogen-activated protein kinase phosphatase-7
PAGE	: polyacryl amide gel electrophoresis
PCR	: polymerase chain reaction
PEG	: polyethylene glycol
R.m.s	: root-mean-square
ROS	: reactive oxygen species
SAD	: single-wavelength anomalous diffraction
SDS	: sodium dodecyl sulfate

SeMet	: selenomethionine
TCEP	: tris(2-carboxyethyl)phosphine
TNF-α	: tumor necrosis factor-α
Tris	: tris(hydroxymethyl)aminomethane
VPS34	: vacuolar sorting protein 34

CURRICULUM VITAE

Kyoung Hoon Kim

PERSONAL

Date of Birth:	Feb 12, 1977
Gender:	Male
Nationality:	Korea, Republic of
Place of Birth:	Seoul, Korea
Nationality:	Korean, Republic of
Marital Status:	Married (Yunjeong Jo) and one son (Tae Gyu)
Address:	Department of Chemistry, College of Natural Sciences Seoul National University, Seoul 151-742, Korea Tel: 82-2-882-3515, Fax: 82-2-889-1568 E-mail: rudgns25@gmail.com

EDUCATION

2003.9 – 2014.2 **Combined M.S. and Ph.D Program in Chemistry (Ph.D.)**

Department of Chemistry, Seoul National University, Korea

"Structural and functional studies of Enhanced intracellular survival proteins from *Mycobacterium tuberculosis* and *Mycobacterium smegmatis*" (Thesis advisor: Prof. Se Won Suh)

1995.3 – 2001.8 **Undergraduate Program in Microbiology (B.S.)**

Department of Microbiology, Pusan National University, Korea

[1997.9 – 1999.11 Obligatory military service]

TEACHING EXPERIENCE

2005 – 2007 **Teaching Assistant**

Department of Chemistry, College of Natural Sciences, Seoul National University, Seoul, Korea

(General Chemistry Laboratory)

RESEARCH EXPERIENCE

2001.8 – 2002.2 Research assistant

Department of Biochemistry, College of Medicine,

	Catholic University of Korea (I carried out biochemical studies in a neuroscience laboratory.)
2002.3 – 2003.7	Research technician Department of Chemistry, College of Natural Sciences, Seoul National University, Korea (I carried out directed evolution in the laboratory of Prof. Se Won Suh.)
2003.8 – present	Ph.D. student Department of Chemistry, College of Natural Sciences, Seoul National University, Korea (I performed structural biology research in the laboratory of Prof. Se Won Suh.)
Many times	Photon Factory, Japan and Pohang Light Source, Korea (I collected and processed X-ray diffraction data.)
2008.3 – 2008.4	School of Biological Sciences, The University of Auckland, New Zealand (I was trained on *Mycobacterium smegmatis* expression system in the laboratory of Prof. Edward N. Baker.)

SKILLS AND TECHNIQUES

Biochemistry	Fast Protein Liquid Chromatography; Electrophoresis; Transfection and mammalian cell culture; Electrophoretic Mobility Shift Assay; Enzyme assay using isotopes; Isothermal Titration Calorimetry; MTT assay; ELISA; Northern blot; Western blot;
Biophysics	Dynamic Light Scattering
Molecular Biology	Cloning; Site-directed mutagenesis; Expression and purification of proteins in bacterial systems; Directed evolution
Structural Biology	Protein crystallography; X-ray data collection and processing; Phasing by MAD/SAD/molecular replacement;

Homology modeling; Model refinement using CNS, CCP4, Phenix

AWARDS AND SCHOLARSHIP

2004.3	National Science and Engineering Graduate Research Scholarship, Korea Student Aid Foundation, Korea
2004.9	Honor Scholarship for high GPA Seoul National University, Korea
2005.3-2006.8	The 1st Seoul Science Fellowship, Seoul Scholarship Foundation, Korea

PUBLICATIONS

First Author publications:

(1) **Kim K. H.,** An D. R., Song J, Yoon J. Y., Kim H. S., Yoon H. J., Im H. N., Kim J, Kim do J., Lee S. J., Kim K. H., Lee H. M., Kim H. J., Jo E. K., Lee J. Y., and Suh S. W. (2012). *Mycobacterium tuberculosis* Eis protein initiates suppression of host immune responses by acetylation of DUSP16/MKP-7. *Proc Natl Acad Sci USA*, **109**, 7729-7734.

(2) Han B. G. *, **Kim K. H.** ***,** Lee S. J. *, Jeong K. C., Cho J. W., Noh K. H., Kim T. W., Kim S. J., Yoon H. J., Suh S. W., Lee S., and Lee B. I. (2012). Helical repeat structure of apoptosis inhibitor 5 reveals protein-protein interaction modules. *J. Biol. Chem.* **287**, 10727-10737. [*These authors contributed equally to this work.]

(3) **Kim K. H.** ***,** Kang W. *, Suh S. W., and Yang J. K. (2011). Crystal structure of FAF1 UBX domain in complex with p97/VCP N domain reveals a conformational change in the conserved FcisP touch-turn motif of UBX domain. *Proteins*, **328**, 505-515. [*These authors contributed equally to this work.]

(4) **Kim K. H.**, Yang J. K., Waldo G. S., Terwilliger T. C., Suh S. W. (2008). From no expression to high-level soluble expression in Escherichia coli by screening a library of the target proteins with randomized N-termini. *Methods Mol. Biol.* **426**, 187-195.

(5) Han S. H. *, Ha J. Y. *, **Kim K. H.** *, Oh S. J., Kim do J., Kang J. Y., Yoon H. J,. Kim S. H., Seo J. H., Kim K. W., and Suh, S. W. (2006). Expression, crystallization and preliminary X-ray crystallographic

analyses of two N-terminal acetyltransferase-related proteins from *Thermoplasma acidophilum*. *Acta Crystallogr*. **F62**, 1127-1130. [*These authors contributed equally to this work.]

(6) **Kim K. H.**, Ahn H. J., Kim D. J., Lee H. H., Ha J. Y., Kim H. K., Yoon H. J., and Suh, S. W. (2005). Expression, crystallization and preliminary X-ray crystallographic analysis of human agmatinase. *Acta Crystallogr.* **F61**, 889-891.

Co-auther publications:

(1) Yoon H. J., **Kim K. H.**, Yang J. K., Suh S. W., Kim H., Jang S. (2013). A docking study of enhanced intracellular survival protein from *Mycobacterium tuberculosis* with human DUSP16/MKP-7. *J Synchrotron Radiat.* **20**, 929-932.

(2) Yoon J. Y., Kim J., Lee S. J., Kim H. S., Im H. N., Yoon H. J., **Kim K. H.**, Kim S. J., Han B. W., and Suh S. W. (2011). Structural and functional characterization of *Helicobacter pylori* DsbG. *FEBS Lett.* **585**, 3862-3867.

(3) Lee S. J., Kim D. J., Kim H. S., Lee B. I., Yoon H.-J., Yoon J. Y., **Kim K. H.**, Jang, J Y., Im H. N., An D. R., Song J.-S., Kim H.-J., and Suh S. W. (2011) Crystal structures of *Pseudomonas aeruginosa* guanidinobutyrase and guanidinopropionase, members of the ureohydrolase superfamily. *J. Struct. Biol.* **175**, 329-338.

(4) Lee S. J. Kim H. S., Kim D. J., Yoon H. J., **Kim K. H.**, Yoon J. Y. and Suh S. W. (2011). Crystal structures of LacD from *Staphylococcus aureus* and LacD.1 from *Streptococcus pyogenes*: insights into substrate specificity and virulence gene regulation. *FEBS Lett.* **585**, 307-312.

(5) Kim D. J., Park K. S., Kim J. H., Yang S. H, Yoon J. Y., Han B. G., Kim H. S., Lee S. J., Jang J. Y., **Kim K. H.**, Kim M. J., Song J. S., Kim H. J., Park C. M., Lee S. K., Lee B. I., and Suh S. W. (2010). *Helicobacter pylori* proinflammatory protein up-regulates NF-kappaB as a cell-translocating Ser/Thr kinase. *Proc. Natl. Acad. Sci. USA* **107**, 21418-21423.

(6) Seo J. H., Cha J. H., Park J. H., Jeong C. H., Park Z. Y., Lee H. S., Oh S. H., Kang J. H., Suh S. W., **Kim K. H.**, Ha J. Y., Han S, H., Kim S. H., Lee J. W., Park J. A., Jeong J. W., Lee K. J., Oh G. T., Lee M. N., Kwon S. W,, Lee S. K., Chun K. H,, Lee S. J., Kim K. W. (2010). Arrest defective 1 autoacetylation is a critical step in its ability to stimulate cancer cell proliferation. *Cancer Res.* **70**, 4422-4432.

(7) Kim H. S., Kim H. L., **Kim K. H.**, Kim D. J., Lee S. J., Yoon J. Y.,

Yoon H. J., Lee H. Y., Park S. B., Kim S.-J., Lee J. Y., and Suh S. W. (2010) Crystal structure of Tpa1 from *Saccharomyces cerevisiae*, a component of the messenger ribonucleoprotein complex. *Nucleic Acids Res.* **38**, 2099-2110

(8) Jang J. Y., Yoon H. J., Yoon J. Y., Kim H. S., Lee S. J., **Kim K. H.**, Kim D. J., Jang S., Han B. G., Lee B. I., and Suh S. W. (2009). Crystal structure of the TNF-α-inducing protein (Tipα) from *Helicobacter pylori*: Insights into its DNA-binding activity. *J. Mol. Biol.* 392(1), 191-197.

(9) Kim D. J., Lee S. J., Kim H. S., **Kim K. H.**, Lee H. H., Yoon H. J., and Suh S. W. (2008). Structural basis of octanoic acid recognition by lipoate-protein ligase B. *Proteins*, **70**, 1620-1625.

(10)Lee H. H., Kim Y.-S., **Kim K. H.**, Heo I., Kim S. K., Kim O., Kim H. K., Yoon J. Y., Kim H. S., Kim D. J., Lee S. J., Yoon H. J., Kim S. J., Lee B. G., Song H. K., Kim V. N., Park C. M., and Suh S. W. (2007). Structural and functional insights into Dom34, a key component of no-go mRNA decay. *Mol. Cell*, **27**, 938-950.

(11) Ha J. Y., Kim H. K., Kim D. J., **Kim K. H.**, Oh S. J., Lee H. H., Yoon H. J., Song H. K., and Suh S. W. (2007). The recombination-associated protein RdgC adopts a novel toroidal architecture for DNA binding. *Nucleic Acids Res.* **35**, 2671-2681.

(12) Ha J. Y., Lee J. H., **Kim K. H.**, Kim D. J., Lee H. H., Kim H.-K., Yoon H.-J., and Suh S. W. (2007). Crystal structure of D-erythronate-4-phosphate dehydrogenase complexed with NAD. *J. Mol. Biol.* **366**, 1294-1304.

(13) Ha J. Y., Min J. Y., Lee S. K., Kim H. S., Kim D. J., **Kim K. H.**, Lee H. H., Kim H. K., Yoon H. J., and Suh S. W. (2006). Crystal structure of 2-nitropropane dioxygenase complexed with FMN and substrate: Identification of the catalytic base. *J. Biol. Chem.* 281(27), 18660-18667.

(14) Ha J. Y., Lee J. H., **Kim K. H.**, Kim D. J., Lee H. H., Kim H.-K., Yoon H.-J., and Suh S. W. (2006). Overexpression, crystallization, and preliminary X-ray crystallographic analysis of erythronate-4-phosphate dehydrogenase from *Pseudomonas aeruginosa*. *Acta Crystallogr.* **F62**, 139-141.

(15) Lee H. H., Yoon J. Y., Kim H. S., Kang J. Y., **Kim K. H.**, Kim D. J., Ha J. Y., Mikami B., Yoon H. J., and Suh S. W. (2006) Crystal structure of a metal ion-bound IS200 transposase. *J. Biol. Chem.* **281**, 4261-4266.

(16) Kim D. J., **Kim, K H.**, Lee H. H., Lee S. J., Ha J. Y., Yoon H. J., and Suh S. W. (2005). Crystal structure of lipoate-protein ligase A bound

with the activated intermediate: Insights into interaction with lipoyl domains. *J. Biol. Chem.* **280**, 38081-38089.

(17) Ahn H. J., **Kim K. H.**, Lee J., Ha J. Y., Lee H. H., Kim D. J., Yoon H. J., Kwon A. R., and Suh S. W. (2004). Crystal structure of agmatinase reveals structural conservation and inhibition mechanism of the ureohydrolase superfamily. *J. Biol. Chem.* **279**, 50505-50513.

(18) Lee J. A., Ahn H. J., Ha J. Y., Shim S. M., **Kim K. H.**, Kim H. K., and Suh S. W. (2004). Crystallization and preliminary X-ray crystallographic analysis of a putative agmatinase from *Deinococcus radiodurans*. *Acta Crystallogr.* **D60**, 1890-1892.

(19) Lee B. I., **Kim K. H.**, Park S. J., Eom S. H., Song H. K., and Suh S. W. (2004). Ring-shaped architecture of RecR: implications for its role in homologous recombinational DNA repair. *EMBO J.* **23**, 2029-2038.

(20) Lee B. I., **Kim K. H.**, Shim S. M., Ha K. S., Yang J. K., Yoon H. J., Ha J. Y., and Suh S. W. (2004). Crystallization and preliminary X-ray crystallographic analysis of the RecR protein from *Deinococcus radiodurans*. *Acta Crystallogr*. **D60**, 379-381.

PRESENTATIONS

2012.7	"*Mycobacterium tuberculosis* Eis protein initiates suppression of host immune responses by acetylation of DUSP16/MKP-7" The 3rd Symposium on Structure and Folding of Disease Related Proteins, Seoul National University, Seoul, Korea (oral presentation in English)
2011.8	"Crystal structures of Eis proteins from *M. tuberculosis and M. smegmatis*" International Union of Crystallography Congress, Madrid, Spain (poster)
2009.10	"Structural Study of Enhanced Intracellular Survival protein, Eis protein from *Mycobacterium tuberculosis*" Asian Crystallographic Association meeting, Beijing, China (poster)
2009.10	

	"Structural and functional studies of Enhanced Intracellular Survival protein, Eis protein from *Mycobacterium tuberc*ulosis" The Second Korea-New Zealand Structural Biology Symposium, Seoul National University, Korea (oral presentation in English)
2008.04	"Overcoming the difficulties of recombinant protein expression in *E. coli* using Green Fluorescence Protein" The First Korea-New Zealand Structural Biology Symposium, University of Auckland, New Zealand (oral presentation in English)
2006.11	"Crystal structure of Ta0058, N-terminal acetyltransferase complex ARD1 subunit related protein, gene product from *Thermoplasma acidophilum*" Korean Crystallographic Association, Seoul National University, Korea (poster)

Structural and Functional Studies of Enhanced Intracellular Survival Proteins from *Mycobacterium tuberculosis* and *Mycobacterium smegmatis*

1. Introduction

Nearly one-third of the world's population is infected with *Mycobacterium tuberculosis*. This pathogenic bacterium causes tuberculosis, which claims the lives of millions of people every year (Dye and Williams, 2010). Tuberculosis has also become a global health issue due to the increased incidences of multidrug-resistant and extensively drug-resistant strains of *M. tuberculosis* (Chiang et al., 2010). This makes a search for targets of new antituberculosis drugs urgent. *M. tuberculosis* is a highly successful human pathogen, surviving and multiplying within the human macrophage cells of the infected people (Dahl et al., 2001). Therefore, treatment of tuberculosis is difficult, requiring many months of taking a combination of antibiotics. *M. tuberculosis* has the ability to persist in the form of a long-term asymptomatic infection, referred to as latent tuberculosis (Lin and Flynn, 2010). Latent tuberculosis becomes activated, when the body's immune system is weakened. As a result, tuberculosis is the major

cause of the deaths among immuno-compromised AIDS patients (Getahun et al., 2010).

In mycobacterial infection, host innate immune responses may play a crucial role in early protection against *M. tuberculosis* infection, leading to establishment of effective adaptive immunity to tuberculosis (Jo et al., 2007). Additionally, mitogen-activated protein kinase (MAPK) pathways are activated by *M. tuberculosis* or its components and play an essential role in the regulation of innate immune signaling during mycobacterial infection (Jo et al., 2007). As intracellular survival of *M. tuberculosis* plays a central role in its pathogenesis (Wei et al., 2000), it is important to understand the survival strategies of this bacterium within macrophages. *M. tuberculosis* has evolved a number of highly effective survival strategies inside the macrophage (Meena and Rajni, 2010). The best characterized survival mechanism of *M. tuberculosis* is the inhibition of phagosomal maturation and autophagy, between which a functional overlap was suggested (Meena and Rajni, 2010; Ehrt and Schnappinger, 2009; Vergne et al., 2004; Vergne I, et al. 2006). Both processes involve several steps including fusion with lysosomes and a number of protein factors such as Beclin 1 and vacuolar sorting protein 34 (VPS34), the class III phosphatidylinositol 3-kinase (Kang et al. 2011). The identification and characterization of mycobacterial

proteins that play a role in facilitating intracellular survival remain a priority for the development of new antituberculosis drugs.

The Rv2416c gene of *M. tuberculosis* H37Rv strain was found to enhance intracellular survival of *Mycobacterium smegmatis* (*M. smegmatis*) in the human macrophage-like cell line U-937 and thus it was designated as *eis* (*e*nhanced *i*ntracellular *s*urvival) (Wei et al., 2000) (Figure 1). The expression of its protein product directly correlated with the enhanced mycobacterial survival in U-937 cells (Wei et al., 2000). The *M. tuberculosis* Eis protein is produced during human tuberculosis infection and is released into the culture medium (Dahl et al., 2001). The sigma factor SigA was shown to bind to the *eis* promoter in the W-Beijing strain of *M. tuberculosis* and the activation of the *M. tuberculosis eis* gene correlated with increased SigA levels and enhanced intracellular survival (Wu et al., 2009). Treatment of T-cells with *M. tuberculosis* Eis inhibited extracellular signal-regulated kinase 1/2 (ERK1/2), JAK pathway, and subsequent production of tumor necrosis factor-α (TNF-α) and interleukin-4 (IL-4) (Lella and Sharma, 2007). *M. tuberculosis* Eis negatively regulated the secretion of IL-10 and TNF-α by primary human monocytes in response to infection with the pathogen (Samuel et al., 2007).

Recently, *M. tuberculosis* Eis was shown to suppress host innate immune defenses by negatively modulating inflammation, autophagy, and cell death in a redox-dependent manner (Shin et al., 2010). The reported data indicate that *M. tuberculosis* Eis plays an essential role in regulating both the early generation of reactive oxygen species (ROS) and inflammatory responses in macrophages (Shin et al., 2010). It was also found that abrogated production of both ROS and proinflammatory cytokines by *M. tuberculosis* Eis depends on its *N*-acetyltransferase domain in the N-terminus (Shin et al., 2010). Enhanced macrophage survival by *M. tuberculosis* Eis was found to occur through the regulation of ROS signaling, which was c-Jun N-terminal kinase (JNK)-dependent but was not p38- or ERK1/2-dependent (Shin et al., 2010). Forced expression of dual-specificity protein phosphatase 16 (DUSP16), also called MAPK phosphatase-7 (MKP-7), suppressed activation of MAPKs in COS-7 cells in the order of selectivity, JNK >> p38 > ERK, suggesting that DUSP16/MKP-7 works as a JNK-specific phosphatase (Masuda et al., 2001).

A bioinformatic analysis predicted that the *M. tuberculosis* Eis protein contains a single acetyltransferase domain of the GCN5-related *N*-acetyltransferase (GNAT) superfamily in the amino terminus (Samuel et al.,

2007). The acetyltransferase domain is predicted to cover residues 9–160 of the 408-residue protein and contains a variant of the characteristic sequence motif (V/I-x-x-x-x-Q/R-x-x-G-x-G/A) for acetyltransferases at positions between 93 and 103 (^{93}VAPTHRRRGLL^{103}) (Samuel et al., 2007; Neuwald and Landsman 1997) (Figure 2). Sequence numbering of *M. tuberculosis* Eis follows the current EXPASY UniProtKB/Swiss-Prot database; six residues ^{1}MPQSDS^{6} at the amino terminus are missing from other databases due to a different translation initiation at Val7. Increased expression of *M. tuberculosis* Eis due to a mutation in the promoter region of the *eis* gene conferred resistance to aminoglycoside kanamycin, a second-line antituberculosis drug (Zaunbrecher et al., 2009). More recently, *M. tuberculosis* Eis was demonstrated to have an unprecedented ability to acetylate multiple amines of many aminoglycosides and its reaction mechanism was proposed on the basis of structural and mutational studies (Chen et al., 2011). However, the fundamental question about the mechanism of enhanced survival of the *M. smegmatis* clone, which contains the extra *M. tuberculosis eis* gene in addition to its own *eis* gene, still remains unanswered. Nonpathogenic *M. smegmatis* contains a homologous

eis gene (M. SMEGMATISEG_3513) that encodes a homolog of *M. tuberculosis* Eis (58% amino acid sequence identity).

M. tuberculosis Eis enhances macrophage survival through the regulation of JNK-dependent ROS signaling (Shin et al., 2010), while DUSP16/MKP-7 works as a JNK-specific phosphatase *in vivo* (Masuda et al., 2001). Acetylation at Lys57 of DUSP1/MKP-1, a phosphatase that inactivates MAPK members by dephosphorylation, increased the interaction of DUSP1/MKP-1 with its substrate p38 MAPK and inhibited innate immune signaling (Cao et al., 2008; Chi and Flavell, 2008).

On the basis of these reports, I spectulated that *M. tuberculosis* Eis may acetylate component(s) of the MAPK signaling pathways, such as MAPK phosphatases, to negatively control autophagy, phagosome maturation, and ROS generation, ultimately leading to the suppression of host immune responses.

To better understand how *M. tuberculosis* Eis enhances mycobacterial survival in macrophages, I have carried out functional and structural characterization of Eis proteins from both *M. tuberculosis* and *M. smegmatis*. I find that both Eis proteins can acetylate aminoglycosides efficiently. I have discovered that Lys55 within the docking domain (also called the kinase interaction motif) of DUSP16/MKP-7 is readily acetylated by *M.*

tuberculosis Eis, but not by *M. smegmatis* Eis. Furthermore, I show that *M. tuberculosis* Eis, but not *M. smegmatis* Eis, inhibits lipopolysaccharide-induced JNK phosphorylation. The observed functional difference against the DUSP16/MKP-7 peptide and JNK can be understood by comparing the structures of the two Eis proteins. The overall monomeric and oligomeric structures of both *M. tuberculosis* Eis and *M. smegmatis* Eis are highly similar to each other. The most notable structural difference between them is the presence of a narrow channel for potential sequence-specific binding of the acetylation target peptide in the active site of *M. tuberculosis* Eis but not in *M. smegmatis* Eis. On the basis of these findings, I propose that *M. tuberculosis* Eis initiates the inhibition of autophagy and phagosome maturation in infected macrophages by N^{ε}-acetylating Lys55 of DUSP16/MKP-7 to suppress host immune responses for intracellular survival of mycobacteria.

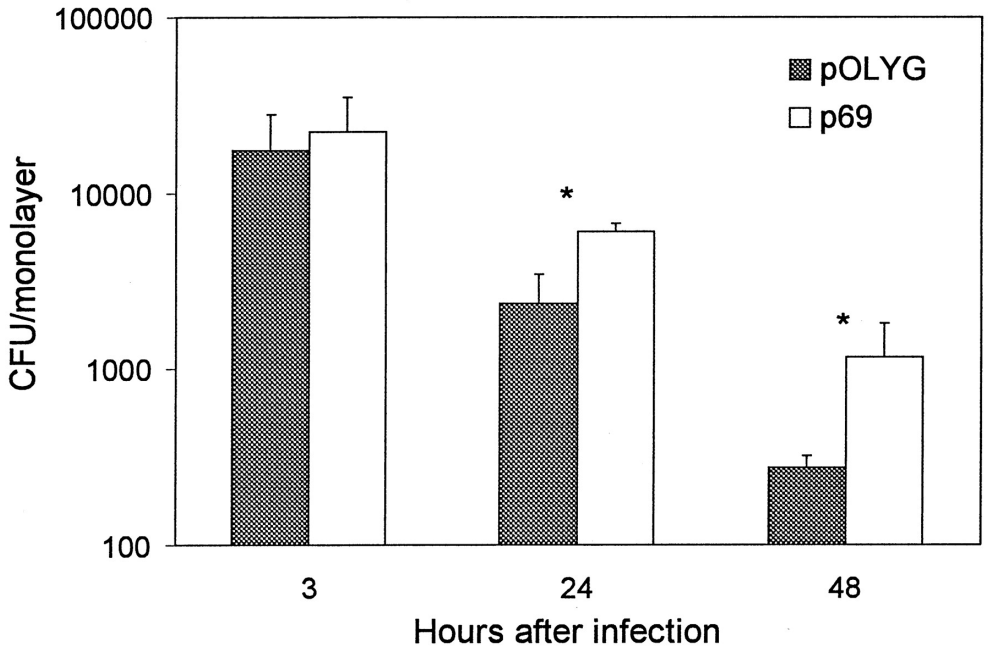


Figure 1. Survival of *M. smegmatis* containing pOLYG (control, shuttle vector) or p69 (pOLYG +*M. tuberculosis* Eis open reading frame) in U-937 cells (Wei et al., 2000)

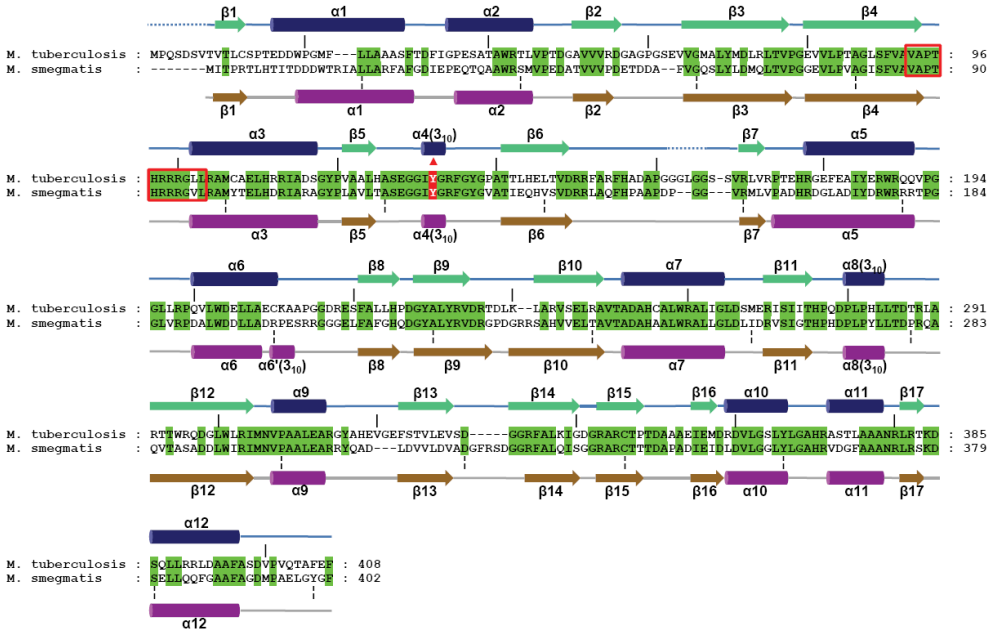


Figure 2. Sequence alignment of *M. tuberculosis* and *M. smegmatis* Eis

Conserved residues are shaded in green. Cylinders denote α- and 3_{10}-helices and arrows denote β-strands, respectively. The P-loop sequence motif is enclosed in a red box. A red triangle above the sequences indicates the conserved tyrosine shaded in red (Tyr132 in *M. tuberculosis* Eis; Tyr126 in *M. smegmatis*). Every 20th residue is marked with a vertical bar and a broken bar above the *M. tuberculosis* Eis sequence and below the *M. smegmatis* Eis sequence, respectively. This figure was drawn with ClustalX (Thompson et al., 1997) and GeneDoc(http://www.psc.edu/biomed/genedoc).

2. Material and methods

2.1. Protein expression, purification, and mutagenesis

The *Rv2416c* gene encoding *M. tuberculosis* Eis was amplified by the polymerase chain reaction using the genomic DNA of *M. tuberculosis* H37Rv as the template. It was inserted into the BamHI/NotI-digested expression vector pET-28b(+) (Novagen), resulting in a twenty-residue hexahistidine-containing tag at its N-terminus. The recombinant *M. tuberculosis* Eis protein was expressed in *E. coli* Rosetta II(DE3)pLysS cells (Invitrogen). The cells were grown at 37°C up to OD_{600} of ~0.5 in Terrific Broth medium that contained 30 μg ml^{-1} kanamycin and the protein expression was induced by 0.5 mM isopropyl-β-D-thiogalactopyranoside (IPTG). The cells were continued to grow at 18°C for 6 h after IPTG induction and were harvested by centrifugation at 4,200 *g* (6,000 rev min^{-1}; Sorvall GSA rotor) for 10 min at 4°C. The cell pellet was resuspended in an ice-cold lysis buffer (5 mM imidazole, 500 mM NaCl, 20 mM Tris-HCl at pH 7.9, 10%(v/v) glycerol, 1 mM phenylmethylsulfonyl fluoride, 0.1 mM tris(2-carboxyethyl)phosphine (TCEP) and was then homogenized with an ultrasonic processor. The crude cell extract was centrifuged at 36,000 *g* (18,000 rev min^{-1}; Hanil Supra 21K rotor) for 60 min at 4°C and the

recombinant protein in the supernatant fraction was purified in three chromatographic steps. The first step was the metal-chelate chromatography on Ni-NTA resin (Qiagen). The protein was eluted with a lysis buffer containing 500 mM imidazole and the eluted sample was diluted five-fold with buffer A (20 mM Tris-HCl at pH 8.5, 0.1 mM TCEP). The diluted sample was applied to a Source 15Q ion-exchange column (GE Healthcare), which was previously equilibrated with buffer A. The protein was eluted with a linear gradient of 0–1.0 M NaCl in buffer A. The next step was gel filtration on a HiLoad 16/60 Superdex-200 prep-grade column (GE Healthcare), employing an elution buffer of 20 mM Tris-HCl at pH 8.5, 0.1 mM TCEP, and 150 mM NaCl. The purified protein was concentrated to 5.0 mg/ml using Centricon YM-10 (Millipore).

For overexpressing the selenomethionine (SeMet)-substituted Eis protein from *M. tuberculosis* in *E. coli* Rosetta(DE3) cells, I used the M9 cell culture medium that contained extra amino acids including SeMet. When the cell culture reached an OD_{600} of 0.6, SeMet was added at 50 mg/L and at the same time the synthesis of Met was repressed by the addition of Phe, Thr, and Lys at 100 mg L^{-1}, and Leu, Ile, Val, and Pro at 50 mg L^{-1} (20). After 15 min, expression of the SeMet-substituted protein was induced with 0.4 mM

IPTG and the culture was grown for additional 20 h at 20°C. The procedure for purifying the SeMet-substituted protein was the same except for the presence of 10 mM dithiothreitol in all buffers used during purification steps besides buffer A.

The recombinant *M. smegmatis* Eis with an N-terminal hexahistidine-containing tag was expressed and purified essentially as above, except the following differences. *E. coli* cells were grown at 37°C for 6 h after IPTG induction and the concentration of Tris-HCl was 50 mM in the lysis buffer and buffer A. The purified *M. smegmatis* Eis protein was concentrated to 43.0 mg/ml.

The gene encoding the human DUSP16/MKP-7 (residues 1–153) fused with a C-terminal hexahistidine-containg tag was cloned into the NdeI/XhoI-digested expression vector pET-21a(+) (Novagen). The recombinant protein was expressed in *E. coli* Rosetta II(DE3) cells (Invitrogen) and purified essentially as above. The full-length DUSP16/MKP-7 was not expressed in *E. coli*.

2.2. Crystallization and X-ray data collection

Crystals *M. tuberculosis* Eis were grown at 24°C using both the hanging drop and sitting drop vapor diffusion methods (Figure 3A). Each hanging or sitting drop was prepared by mixing equal volumes (1 μl each) of the *M. tuberculosis* Eis protein solution (at 5.0 mg ml^{-1} concentration in 20 mM Tris-HCl at pH 8.5, 0.1 mM TCEP, 150 mM NaCl) and the reservoir solution. The reservoir solution was 480 μl for the hanging drop experiments and 75 μl for the sitting drop experiments. $P2_1$ crystals of the SeMet-substituted *M. tuberculosis* Eis in the acetyl CoA-bound form were grown in about two days using a reservoir solution of 12%(w/v) PEG 3350 and 300 mM ammonium citrate at pH 7.0. Acetyl CoA was not intentionally added to the protein solution during the crystallization step, because it caused precipitation. Acetyl CoA was apparently picked up by the recombinant protein in *E. coli* and remained bound to the protein throughout purification and crystallization steps. H32 crystals of the native *M. tuberculosis* Eis in the apo form were obtained in about two days using the same reservoir solution of 12%(w/v) PEG 3350 and 300 mM ammonium citrate at pH 7.0. Apparently, the purified protein sample contained both acetyl CoA-bound and unbound Eis.

$P2_12_12_1$ crystals of *M. smegmatis* Eis were grown at 24°C using sitting drop vapor diffusion method (Figure 3B). Each sitting drop was prepared by mixing equal volumes (1 μl each) of the *M. smegmatis* Eis protein solution (at 43.0 mg ml^{-1} concentration in 20 mM Tris-HCl at pH 8.5, 0.1 mM TCEP, 150 mM NaCl) and the reservoir solution. The reservoir solution (75 μl) contained 100 mM sodium chloride, 100 mM sodium citrate at pH 5.6, and 12%(w/v) PEG 4000. CoA was not intentionally added during the crystallization step but was apparently picked up by the recombinant protein in *E. coli* and remained bound to the protein throughout purification and crystallization steps.

Single-wavelength anomalous diffraction (SAD) data were collected at 100 K from a crystal of the SeMet-substituted *M. tuberculosis* Eis protein in the acetyl CoA-bound form at the peak wavelength on a Quantum 315 CCD detector (Area Detector Systems Corporation, Poway, California) at the BL-5A experimental station of Photon Factory, Japan. The crystal was flash-frozen using a cryo-protectant, which was the reservoir solution supplemented with 30%(v/v) glycerol. Raw data were processed and scaled using the program HKL2000 and SCALEPACK (Otwinowski and Minor, 1997). The space group was $P2_1$, with unit cell parameters of a = 108.13 Å,b

= 150.19 Å, c = 184.38 Å, β = 103.05° (Table 1). Assuming two Eis hexamers in the asymmetric unit, the Matthews coefficient is 2.55 $Å^3$ Da^{-1} and the solvent content of 51.8%.

The apo crystal of the native *M. tuberculosis* Eis was flash-frozen using a cryo-protectant, which was the reservoir solution supplemented with 30%(v/v) glycerol. X-ray diffraction data were collected on a Quantum 315 CCD detector at the BL38B1 experimental station of SPring-8, Japan. The space group was H32, with unit cell parameters of a = b = 174.57 Å, c = 122.83 Å, α = β = 90°, and γ = 120° (Table 1). The presence of one Eis monomer in the asymmetric unit gives the Matthews coefficient of 3.75 $Å^3$ Da^{-1} and the solvent content of 67.2%.

A crystal of *M. smegmatis* Eis was transferred to a cryoprotectant solution which contained 30%(v/v) glycerol in the reservoir solution. X-ray diffraction data were collected on a Quantum 315 CCD detector at the BL-4A experimental station of Pohang Light Source, Korea. The space group was $P2_12_12_1$, with unit cell parameters of a = 107.74 Å, b = 126.52 Å, c = 238.80 Å, α = β = γ = 90° (Table 1). The presence of one hexamer in the asymmetric unit gives the Matthews coefficient of 2.94 $Å^3$ Da^{-1} and the solvent content of 58.2%.

2.3. Structure solution and refinement

Using the SAD data of the SeMet-labeled *M. tuberculosis* Eis crystal in the acetyl CoA-bound form, 72 of the 84 expected selenium atoms in twelve Eis monomers were located with the SnB program package (Miller et al., 1994), except those of the amino terminal methionines. The phases were calculated by the program SOLVE (Terwilliger and Berendzen, 1999) and were improved through non-crystallographic symmetry averaging, solvent flattening, and histogram matching with the program RESOLVE (Terwilliger, 2000). An initial model was built manually using the program Coot (Emsley and Cowtan, 2004) and the model was refined with the program packages CCP4 (Collaborative Computational Project, Number 4, 1994), CNS (Brünger et al., 1998), and PHENIX (Adams et al., 2010). Subsequently, this model was used as a probe to solve the structures of *M. tuberculosis* Eis in the apo form and *M. smegmatis* Eis (in the CoA-bound form) by the molecular replacement method using the program PHASER (McCoy et al., 2007). The refined models were evaluated using MolProbity (Chen et al., 2010).

2.4. Spectrophotometric acetylation assay of aminoglycosides and steady-state kinetic measurements

The spectrophotometric assay of acetyltransferase activity was performed by coupling the free thiol group of CoA, the reaction product, to 4,4'-dithiodipyridine to produce thiopyridone, which can be monitored by an increase in absorbance at 324 nm (ε_{324} = 19,800 $M^{-1}cm^{-1}$) (Chen et al., 2011; Hegde et al., 2001). The reaction was initiated by adding the Eis protein solution (final concentration 0.5 μM) to the reaction mixture (total 100 μL) containing acetyl CoA (0.1 mM or 1 mM), aminoglycoside (0.1 mM), 4,4'-dithiodipyridine (2 mM), and Tris-HCl at pH 8.0 (50 mM) at 25°C. Reactions were monitored by taking absorbance readings at 1 min interval for the first 60 min and at 5 min interval for additional 120 min in a 96-well plate format. The kinetic parameters for acetylation of aminoglycosides by Eis were determined at 25°C in reaction mixtures (total 100 μL) containing a fixed concentration of acetyl CoA (0.1 mM), varied concentrations of aminoglycoside (0, 5, 10, 20, 50, 100, 250, 500 μM), 4,4'-dithiodipyridine (2 mM), Tris-HCl at pH 8.0 (50 mM), and Eis (0.25 μM). Reactions were initiated by the addition of acetyl CoA and were carried out in duplicate. Net acetylation was monitored by measuring the absorbance at 324 nm as a function of time for each concentration of aminoglycoside. The kinetic parameters, K_m and k_{cat}, were determined from the linear parts of these time courses by using Lineweaver-Burk plots.

2.5. *In vitro* acetylation of synthetic peptides and mass spectrometry

The Eis protein (4 nmol) and the peptide (4 nmol) were incubated at 37°C for 30 min in a buffer (total 30 μl) containing 50 mM Tris–HCl at pH 8.0, 10%(v/v) glycerol, 1 mM DTT, 1 mM EDTA, 10 mM sodium butyrate, and 1 mM acetyl CoA. As a negative control, I also monitored acetylation in the presence of acetyl CoA only. The MALDI-TOF mass spectrometry (MS) analysis was done as in (Shin and Kim, 2011). Peptide fragmentation and MALDI MS/MS analysis were carried out in the positive ion mode using the LIFT-TOF/TOF technique on an Autoflex II MALDI-TOF/TOF mass spectrometer (Bruker Daltonics). Data were analyzed using Flex Analysis 2.4 and Bio-Tools 3.0 software (Bruker Daltonics).

2.6. *In vitro* protein *N*-acetylation assay

To assay the protein *N*-acetyltransferase activity, His-tagged Eis protein (400 ng) and human DUSP16/MKP-7 (residues 1–153) (500 ng) were incubated in the buffer (50 mM Tris-HCl at pH 8.0, 0.1 mM EDTA, 1 mM dithiothreitol, 10 mM sodium butyrate, 10%(v/v) glycerol) and 3 μl of [^{14}C]-labeled acetyl CoA (0.05 mCi ml^{-1}, 53 mCi mmol^{-1}) was added in a total volume of 30 μl. After incubation at 37°C for up to 120 min, the reaction

products were analyzed by SDS-PAGE and the protein bands were stained by Coomassie brilliant blue. Acetylated proteins were visualized using a Bioimage analyzer (BAS-1500, Fuji Film, Japan).

2.7. Cell culture and Western blotting

Bone marrow-derived macrophages (BMDMs) were isolated and then differentiated by growth for 5–7 days in a medium containing macrophage-colony stimulating factor (25 μg mL^{-1}; R&D Systems), as described previously (Shin et al., 2010). Treated BMDMs were processed for analysis by Western blotting as described previously (Shin et al., 2010). For Western blot analysis, primary antibodies were diluted 1:1000. Membranes were developed using a chemiluminescent reagent (Amersham ECL; GE Healthcare) and subsequently exposed to film. Data obtained from independent experiments (presented as mean ± SD) were analyzed by the paired Student's *t*-test. A p value < 0.05 was deemed to indicate statistical significance.

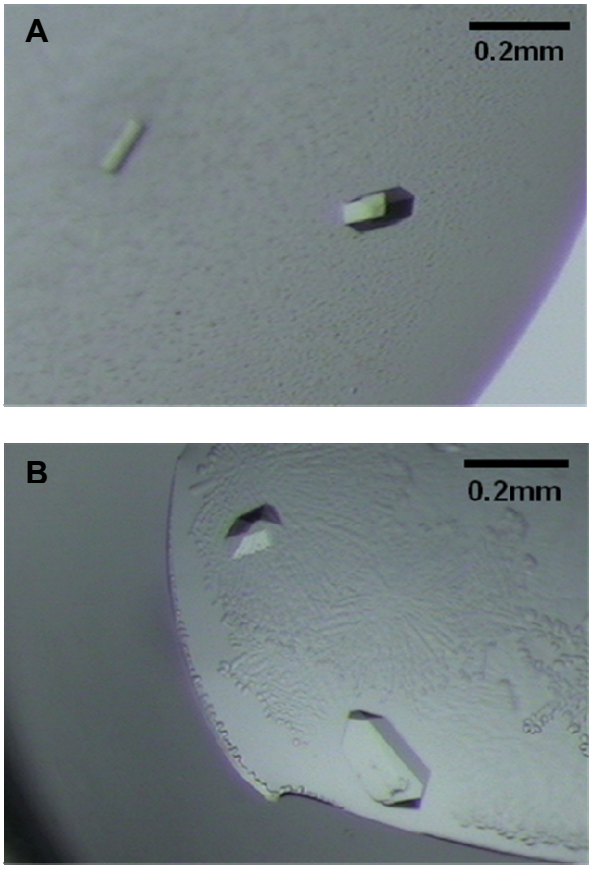


Figure 3. Crystals of Eis from *M. tuberculosis and M. smegmatis*

(A) Crystals of Eis from *M. tuberculosis*. They belong to a space group $P2_1$. Their approximate size is 0.2 mm × 0.05 mm × 0.05 mm.

(B) Crystals of Eis from *M. smegmatis*. They belong to a space group $P2_12_12_1$. Their approximate size is 0.2 mm × 0.1 mm × 0.05 mm.

Table 1. Data collection, phasing, and refinement statistics

A. Data collection and phasing statistics			
Protein	*M. tuberculosis* Eis		*M. smegmaits* Eis
Data set	Acetyl CoA (SeMet SAD)	Apo (Native)	CoA (Native)
Resolution range (Å)	30.0–2.80 (2.85–2.80)[a]	30.0–2.46 (2.55-2.46)[a]	30.0–2.03 (2.06–2.03)[a]
X-ray source	Photon Factory BL-5A	SPring-8 BL-38B1	PLS BL-4A
X-ray wavelength (Å)	0.9792 (peak)	1.0000	1.0000
Space group	$P2_1$	H32	$P2_12_12_1$
a, b, c (Å)	108.13, 150.19, 184.38	174.57, 174.57, 122.83	107.74, 126.52, 238.80
α, β, γ (°)	90, 103.05, 90	90, 90, 120	90, 90, 90
Total/unique reflections	977,419 / 269,936[b]	418,719 / 25,478	1,110,387 / 203,855
Completeness (%)	97.1 (95.6)[a]	99.9 (99.1)[a]	96.3 (96.0)[a]
Mean (I/σ_I)	19.3 (3.88)[a]	37.2 (5.83)[a]	36.9 (4.36)[a]
R_{merge} (%)[c]	9.7 (42.7)[a]	11.2 (48.7)[a]	6.0 (39.6)[a]
f' / f'' (e^-)	–8.15 / 3.84		
Figure of merit[d] for SAD phasing (20.0–2.80 Å): 0.50 / 0.72 (before / after density modification)			

B. Refinement statistics			
Resolution range (Å)	30.0–2.80	20.0–2.50	20.0–2.03
R_{work} / R_{free} (%)[e]	19.2 / 24.8	18.3 / 21.6	18.1 / 21.7
No. of residues / mean B-factor ($Å^2$)	4,752 / 40.1	396 / 32.7	2,412 / 31.7
No. of waters / mean B-factor ($Å^2$)	489 / 29.5	137 / 34.1	2,007 / 35.9
No. of ligands / mean B-factor ($Å^2$)	12 / 53.5	0	6 / 32.3
R.m.s.deviation			
Bond lengths (Å)	0.008	0.008	0.007
Bond angles (°)	1.23	1.11	1.19
Ramachandran plot			
Most favorable (%)	95.7	97.7	98.5
Allowed (%)	4.2	2.3	1.5
Generously allowed (%)	0.1	0.0	0.0

Footnotes for Table 1

[a] Values in parentheses refer to the highest resolution shell.

[b] Values obtained by treating Friedel pairs as separate observations for the SAD data.

[c] $R_{merge} = \Sigma_h\Sigma_i \mid I(h)_i - \langle I(h) \rangle \mid / \Sigma_h\Sigma_i I(h)_i$, where I(h) is the intensity of reflection h, Σ_h is the sum over all reflections, and Σ_i is the sum over i measurements of reflection h.

[d] Figure of merit = $\langle \mid \Sigma P(\alpha)e^{i\alpha} / \Sigma P(\alpha) \mid \rangle$, where α is the phase angle and $P(\alpha)$ is the phase probability distribution.

[e] $R = \Sigma \mid |F_{obs}| - |F_{calc}| \mid / \Sigma |F_{obs}|$, where R_{free} is calculated for a randomly chosen 10% of reflections, which were not used for structure refinement, and R_{work} is calculated for the remaining reflections.

3. Results and discussion

3.1. Overall structure and structural similarity search

The Eis monomer of both *M. tuberculosis* and *M. smegmatis* is tripartite, consisting of three "structural" domains (Figure 4). "Structural" domain 1 (residues 1–142 and 298–305 in *M. tuberculosis*; residues 1–136 and 290–297 in *M. smegmatis*) shows the GNAT fold, as predicted by sequence analysis. Unexpectedly, "structural" domain 2 (residues 143–297 in *M. tuberculosis*; residues 137–289 in *M. smegmatis*) shares the same GNAT fold, despite no significant sequence identity with "structural" domain 1 (6.7% identity in *M. tuberculosis*; 4.8% in *M. smegmatis*). "Structural" domain 3 (residues 306–408 in *M. tuberculosis*; residues 298–402 in *M. smegmatis*) resembles the sterol carrier protein-2 (SCP-2) domain, which was also not predictable from the amino acid sequence due to no detectable sequence identity.

"Structural" domain 1, folded into the GNAT structure, has a six-stranded β sheet flanked by two α helices (α1 and α2) on one side, and by one α helix (α3) and a 3_{10} helix (α4) on the other side. The central β sheet is arranged in the order β1↑-β2↓-β3↑-β4↓-β5↓-β12(C-terminal part)↑. In this

domain, the N-terminal β1 strand is "structurally" preceded by the C-terminal part of β12 strand and is followed by two helices (α1 and α2), three antiparallel β strands (β2-β3-β4), a "signature" central helix (α3) (1), a fifth β strand (β5), a 3_{10} helix (α4), and a sixth β strand (β12). The N-terminal part (residues 291–297 in *M. tuberculosis*; residues 283–289 in *M. smegmatis*) of the long strand β12 belongs to the central β sheet of "structural" domain 2, while its C-terminal portion (residues 298–304 in *M. tuberculosis*; residues 290–296 in *M. smegmatis*) forms part of the central β sheet of "structural" domain 1. The long strand β12 serves as a linker between "structural: domains 2 and 3. "Structural" domain 1 of *M. tuberculosis* Eis contains a β bulge in strand β4 at Ser89 and Phe90, and an extended P-loop between β4 and α3 (Figure 5). "Structural" domain 1 of *M. smegmatis* Eis contains a β bulge in strand β4 at Ser83 and Phe84, and an extended P-loop between β4 and α3. They are conserved features of other GNAT superfamily members (Vetting et al., 2005). The β bulge splits the central β sheet into two arms, resulting in the V-shaped acetyl CoA binding site between β4 and β5. The P-loop plays a role in acetyl CoA (or CoA) recognition.

Unexpectedly, "structural" domain 2 (residues 143–297 in *M. tuberculosis*; residues 137–289 in *M. smegmatis*) is also folded into the GNAT structure. Between "structural" domains 1 and 2, the root-mean-

square (r.m.s.) deviation is 3.0 Å for 150 Cα atom pairs in acetyl CoA-bound *M. tuberculosis* and 2.6 Å for 144 Cα atom pairs in *M. smegmatis*. "Structural" domain 2 contains a seven-stranded β sheet packed on one side by a pair of helices (α5 and α6) and on the other side by another pair of helices (α7 and α8). The strands of the β sheet are arranged in the order β7↑-β8↓-β9↑-β10↓-β11↓-β6↑-β12 (N-terminal part)↓. In "structural" domain 2 of *M. tuberculosis* Eis, a β bulge is present at Ser245 and Glu246 in strand β10, whereas the P-1oop located between β10 and α7 is highly truncated (Figure 5). In "structural" domain 2 of *M. smegmatis* Eis, a β bulge is present at Val237 and Glu238 in strand β10, whereas the P-1oop located between β10 and α7 is highly truncated. Consistent with this observation, acetyl CoA (or CoA) is bound to domain 1 only but not to domain 2 in ligand-bound structures. Furthermore, a conserved tyrosine residue (Tyr132 in *M. tuberculosis*; Tyr126 in *M. smegmatis*) in "structural" domain 1, which is predicted to be important for the acetyltransferase activity (Vetting et al., 2005), is replaced by a leucine residue (Leu281 in *M. tuberculosis*; Leu272 in *M. smegmatis*) in "structural" domain 2 (Figure 2). This suggests that only "structural" domain 1 is catalytically active and "structural" domain 2 may serve a different noncatalytic function such as substrate binding.

"Structural" domain 3 (residues 306–408 in *M. tuberculosis*; residues 298–402 in *M. smegmatis*) contains a five-stranded β sheet in the order β15↓-β14↑-β13↓-β16↓-β17↑ (Figure 4). One face of the β-sheet is covered by four α helices (α9, α10, α11, and α12), whereas the other side is not covered by α helices (Figure 4). It is structurally related to SCP-2 domains (Stolowich et al., 2002). SCP-2, also called nonspecific lipid transfer protein, is a nonspecific lipid-binding protein expressed ubiquitously in most organisms. In vertebrates, SCP-2 serves not only in cholesterol and phospholipid transfer, but also in regulating multiple lipid signaling pathways in lipid raft/caveolae microdomains of the plasma membrane (Stolowich et al., 2002). In insects, SCP-2 is involved in cholesterol uptake in the midgut in both larval and adult mosquito (Singarapu et al., 2010). Structures of human and mosquito SCP-2 have been reported, both in the apo form and in the ligand-bound forms (Singarapu et al., 2010; García et al., 2000; Haapalainen et al., 2001; Haapalainen et al., 2001). Crystal structures of the putative SCP-2 (TT1886) from *Thermus thermophilus* HB8 have been determined in both open and closed conformations (Goroncy et al., 2010). The most prominent structural feature of SCP-2 proteins is the presence of a hydrophobic cavity for lipid binding (Stolowich et al., 2002). Cavity volumes are 1,100 $Å^3$ for human SCP-2 (PDB 1IKT), 750 $Å^3$ for mosquito

SCP-2 (PDB 1PZ4), and 470/570 $Å^3$ for *T. thermophilus* HB8 SCP-2 like protein in the open/closed conformations (PDB 2CX7 chain A/B) (Dundas et al., 2006). "Structural" domain 3 of *M. tuberculosis* Eis has a much smaller cavity volume of 250 and 240 $Å^3$ for acetyl CoA-bound and apo structure of *M. tuberculosis* Eis, whereas that of CoA-bound *M. smegmatis* Eis has no detectable cavity.

A structural similarity search with *M. tuberculosis* Eis monomer (chain A in the acetyl CoA-bound structure) using the program DALI (Holm and Rosenström, 2010) reveals that members of the GNAT superfamily give the highest Z-scores. Excluding Eis, they are (i) GNAT family acetyltransferase from *A. variabilis* ATCC 29413 (PDB code 2OZG, Z score = 39.6, r.m.s. Cα deviation = 2.5 Å, sequence identity = 22% for 371 residues), (ii) GNAT family acetyltransferase from *E. faecalis* (PDB code 2I00, Z score = 37.9, r.m.s. Cα deviation = 2.4 Å, sequence identity = 17% for 364 residues), (iii) a hypothetical protein from *E. faecalis* V583 (PDB code 2HV2, Z score = 37.0, r.m.s. Cα deviation = 2.7 Å, and a sequence identity = 19% for 369 residues), and (iv) GNAT family acetyltransferase from *Bacillus anthracis* (PDB code 3N7Z, Z score = 36.4, r.m.s. Cα deviation = 2.4 Å, sequence identity = 17% for 362 residues). When I further elaborated the structural similarity search using each of "structural" domain 1, 2, or 3, the results

were similar to that obtained using the whole monomer. However, these proteins have not been well characterized functionally.

Below the above four best matches, better characterized proteins that structurally resemble *M. tuberculosis* Eis are as follows. For "structural" domain 1 (chain A, acetyl CoA-bound model), the GNAT domains of *M. tuberculosis* aminoglycoside 2'-*N*-acetyltransferase (AAC; PDB ID 1M4G) (Vetting et al., 2002), *Ovis aries* serotonin *N*-acetyltransferase (PDB ID code 1L0C) (Scheibner et al., 2002), human p300/CBP-associating factor (PCAF; PDB ID 1CM0) (Clements et al., 1999), and *Salmonella typhimurium* N^{α}-acetyltransferase RimL (PDB ID code 1S7L) (Vetting et al., 2005) show r.m.s. differences of 1.8, 1.8, 2.0, and 2.8 Å for 143 Cα atom pairs, respectively. For "structural" domain 2, the GNAT domains of AAC, serotonin *N*-acetyltransferase, PCAF, and RimL show r.m.s. differences of 2.9, 2.5, 2.8, and 3.1 Å for 150 Cα atom pairs, respectively. This result indicates that "structural" domain 2 of *M. tuberculosis* Eis is structurally more divergent from these GNAT enzymes than "structural" domain 1. For "structural" domain 3, SCP-2 proteins are structurally similar. For example, between "structural" domain 3 and the mosquito SCP-2 (PDB code 1PZ4) (Dyer et al., 2003), the r.m.s. deviation is 2.6 Å for 84 equivalent Cα

positions (residues 306–408 of *M. tuberculosis* Eis), with a Z-score of 8.1, and a sequence identity of 17%.

Interestingly, a putative peroxisome targeting signal type 2 (PTS2) is present in "structural" domain 3 of *M. tuberculosis* Eis as ^{380}RLRTKDSQL^{388} (Figure 2). The generally accepted PTS2 sequence is (R/K)-(L/V/I)-xxxxx-(H/Q)-(L/A) (Lazarow, 2006). The crystal structures of two PTS2s in fructose-1,6-bisphosphate aldolases from *Trypanosoma brucei* and *Leishmania mexicana* are virtually identical (Lazarow, 2006). Nine Cα atoms of the putative PTS2 sequence in "structural" domain 3 of *M. tuberculosis* Eis (chain A, acetyl CoA-bound model) overlap with the *T. brucei* (or *L. mexicana*) PTS2 with an r.m.s. deviation of 2.4 Å. The highly conserved four residues are all located on the surface, thus allowing them to possibly interact with the import receptor Pex7p. If the putative PTS2 in "structural" domain 3 of *M. tuberculosis* Eis is functional, Eis may be translocated into the peroxisomal matrix to disturb the normal peroxisome function and may influence the production of hydrogen peroxide, which acts as a cell-survival signaling molecule (Groeger et al., 2009). It is also interesting to note that a conserved residue (His or Gln) at position no. 8 is replaced by Glu in the corresponding sequence (RLRSKDSEL) of *M.*

smegmatis Eis domain 3. Basic amino acids were not tolerated at position no. 8 (Lazarow, 2006). Replacement of Gln by Glu at position no. 8 in the PTS2-like sequence of *M. tuberculosis* Eis alters the surface charge property (Figure 6). All the known functional PTS2s are located at or near the amino terminus (Lazarow, 2006). Even though the PTS2-like sequence of *M. tuberculosis* Eis is located near the carboxyl terminus, further experimental testing of potential new PTS2s is required to determine whether they are functional or not (Lazarow, 2006). It is worth noting that human and mosquito SCP-2 proteins contain a PTS1 sequence (A/S-K-L) at the carboxyl terminus (Lensink et al., 2002), while *T. thermophilus* HB8 SCP-2 like protein contains neither a PTS1-like sequence nor a PTS2-like sequence (Goroncy et al., 2010).

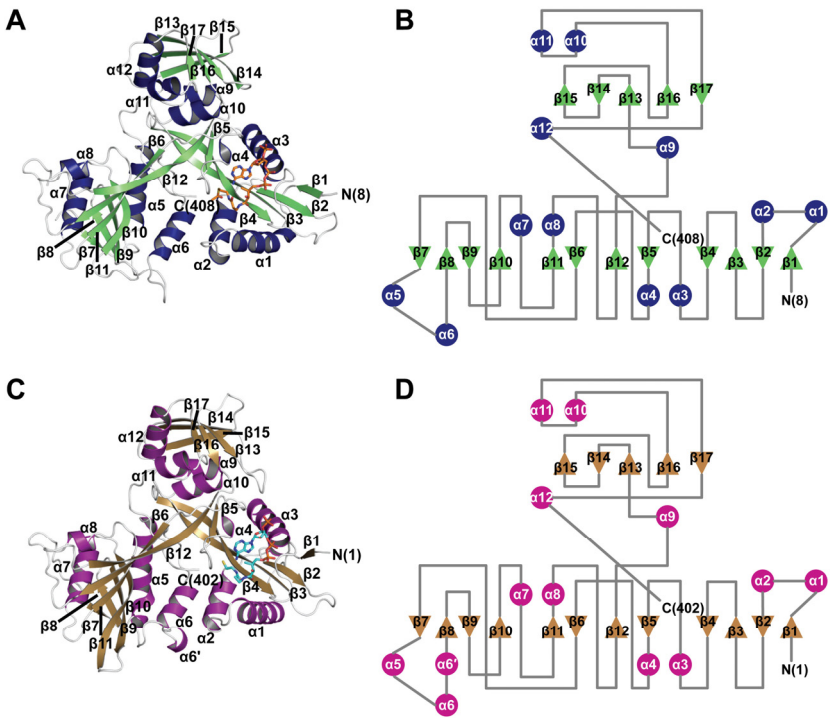


Figure 4. Structures of *M. tuberculosis* and *M. smegmatis* Eis

(A) Cartoon representation of the *M. tuberculosis* Eis monomeric structure. Helices and strands are colored in light blue and lime, respectively. Acetyl CoA is drawn in an orange stick model. (B) Topology diagram of *M. tuberculosis* Eis with helices and strands represented with circles (in blue) and triangles (in lime), respectively. (C) Cartoon representation of the *M. smegmatis* Eis monomeric structure. Helices and strands are colored in purple and orange, respectively. CoA is drawn in a cyan stick model. (D) Topology diagram of *M. smegmatis* Eis with helices and strands represented with circles (in purple) and triangles (in orange), respectively.

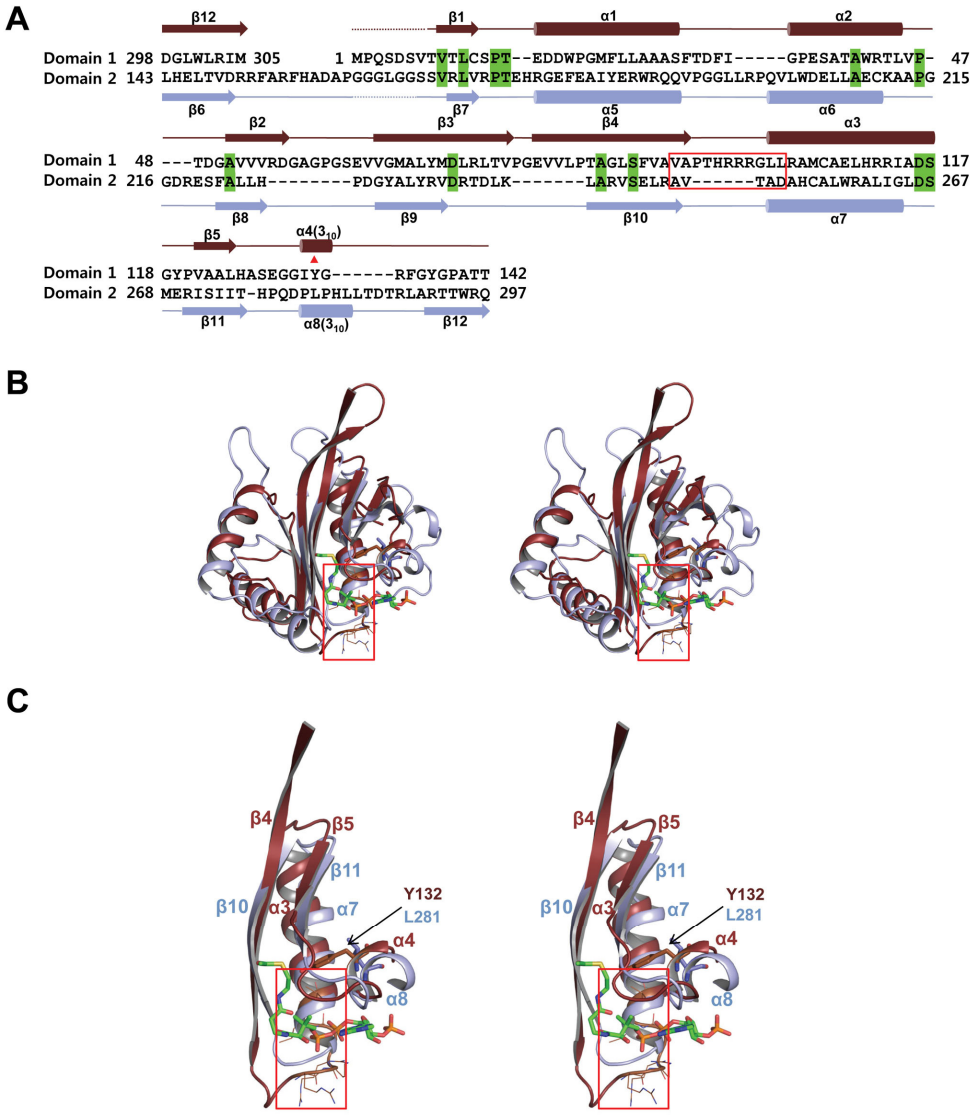


Figure 5. Sequence alignment and structural superposition of domains 1 and 2 of *M. tuberculosis* Eis

(A) Sequence alignment of *M. tuberculosis* Eis "structural" domains 1 and 2. Conserved residues are shaded in green. Cylinders above and below the sequences denote helices and arrows β-strands. Dotted lines represent disordered residues. A red triangle above the sequences indicates Tyr132. The P-loop sequence motif is enclosed in a red box. The N- and C-terminal portions of the long strand β12 are located at the edge of the β-sheets in domains 2 and 1, respectively. (B) Stereo view of the superposition of "structural" domains 1 and 2 in *M. tuberculosis* Eis. "Structural" domains 1 and 2 are colored in ruby and light blue, respectively. Acetyl CoA bound to domain 1 is shown in a stick model. A red box indicates the loop between β4-α3 in "structural" domain 1 and the highly truncated loop between β10-α7 in "structural" domain 2. The P-loops are located in the bottom region of this red box. (C) Close-up, stereo view of acetyl CoA binding to "structural" domain 1. β4/α3/β5/α4 in "structural" domain 1 and β10/α7/β11/α8 in "structural" domain 2 are indicated. A red box indicates the loop between β4-α3 in "structural" domain 1 and the highly truncated loop between β10-α7 in "structural" domain 2. The P-loops are located in the bottom region of this red box. The side chains of Tyr132 in "structural" domain 1 and Leu281 in "structural" domain 2 are indicated by an arrow.

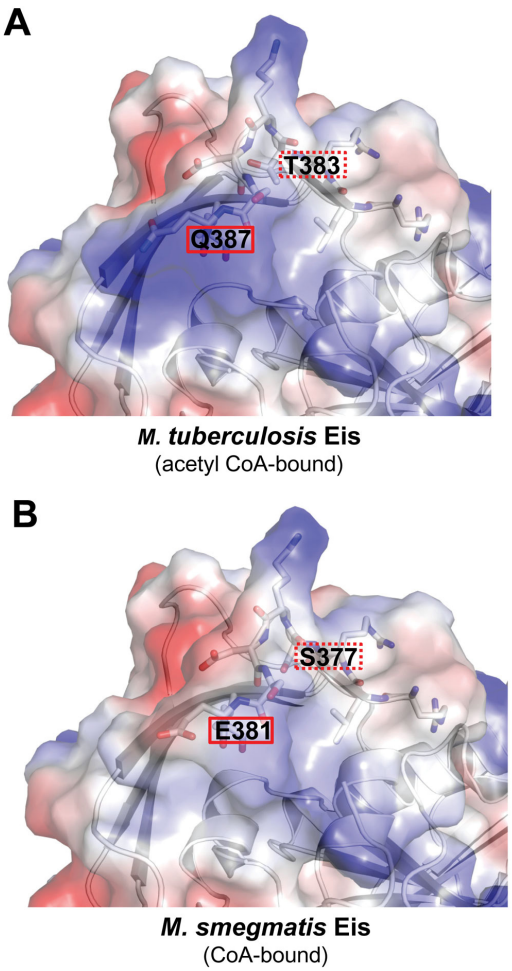


Figure 6. PTS2 region in Eis domain 3

(A) Electrostatic potential at the surface of the PTS2 region in *M. tuberculosis* Eis "structural" domain 3. Blue and red colors correspond to positive and negative potentials, respectively. (B) Electrostatic potential at the surface of the PTS2 region in *M. smegmatis* Eis "structural" domain 3.

Blue and red colors correspond to positive and negative potentials, respectively.

3.2. Quaternary structure of Eis proteins

In all three Eis structures, six monomers are assembled into a homohexamer with point group symmetry 32 (Figure 7). The hexameric association leads to the burial of a total of 27,100, 27,200, and 26,700 $Å^2$ of the solvent-accessible surface area per hexamer for the acetyl CoA-bound and apo structures of *M. tuberculosis* Eis and for CoA-bound *M. smegmatis* Eis, respectively. Extensive interactions across the inter-subunit interfaces in the Eis hexamer suggest that the hexameric structure of Eis observed in the crystal is likely to exist in solution and is biologically relevant.

A hexamer of Eis can be described as a trimer of dimers. The dimer interface buries 5,000, 5,100, and 5,040 $Å^2$ per monomer of solvent-accessible surface area in the acetyl CoA-bound and apo structures of *M. tuberculosis* Eis and for CoA-bound *M. smegmatis* Eis, respectively. "Structural" domains 2 and 3 are primarily involved in the dimer interface. In the *M. tuberculosis* Eis structure, this interface is mainly contributed by residues Gln278–Asp298 from helix α8, strand β12, and the following loop of "structural" domain 2, and by residues Arg371–Arg390 from helices α11 and α12 of "structural" domain 3. When I view the hexamer along its three-fold axis, two triangles representing the two layers of a trimeric unit make a

30° angle. “Structural” domains 1 and 2 alternate within a trimeric unit, whereas “structural” domain 3 is located at the periphery of the triangular-shaped trimeric unit. “Structural” domain 2 is located closer to the interface between two layers of the trimeric unit and two such “structural” domains from different layers make extensive contacts with each other, whereas “structural” domain 1 plays an insignificant role in the contact between the two layers of the trimeric unit. Interestingly, one side of the β-sheet in “structural” domain 3 that is not covered by helices is exposed to the bulk solvent in the hexamer structure.

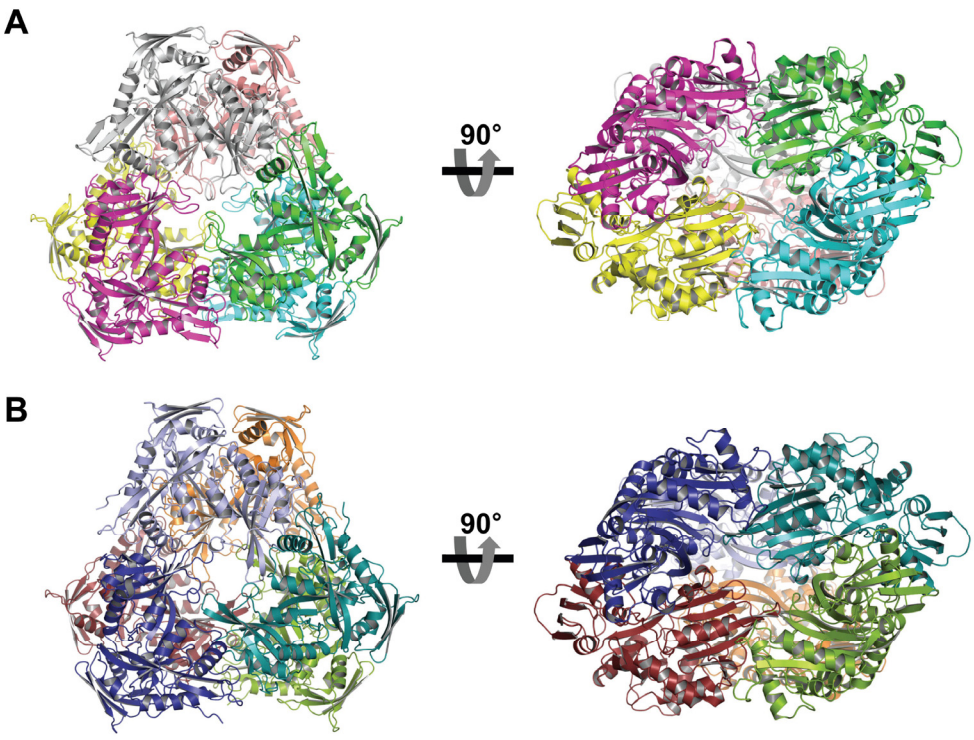


Figure 7. Hexameric arrangement of Eis proteins

(A) Cartoon representation of the *M. tuberculosis* Eis hexameric structure. Two orthogonal views differing by 90° are shown. Individual monomers in the hexamer are colored in gray, green, magenta, pink, cyan, and yellow. (B) Cartoon representation of the *M. smegmatis* Eis hexameric structure. Two orthogonal views differing by 90 are shown. Individual monomers in the hexamer are colored in light blue, teal, blue, light orange, lime, and brown.

3.3. Ligand binding at the active site

The ligand (acetyl CoA or CoA) is bound to "structural" domain 1 but not to "structural" domain 2 in both *M. tuberculosis* and *M. smegmatis* Eis structures, despite sharing the common GNAT fold. The protein-ligand interactions in both Eis structures are highly similar to each other and are also similar to those that have been observed in a number of other GNAT family members. Acetyl CoA (or CoA) is bound between the splayed strands β4 and β5 of "structural" domain 1. The P-loop between the strand β4 and helix α3 coordinates the pyrophosphate group of acetyl CoA (or CoA) (Vetting et al., 2005). The sequences Arg98-Arg99-Arg100-Gly101-Leu102-Leu103 in *M. tuberculosis* Eis and Arg92-Arg93-Arg94-Gly95-Val96-Leu97 in *M. smegmatis* Eis nearly conform to the P-loop motif [(R/Q)-X-X-G-X-(G/A)] for acetyl CoA recognition (Neuwald and Landsman, 1997; Shi et al., 2008). The P-loop is one of the conserved sequence motifs and a conserved structural feature of the GNAT superfamily (Vetting et al., 2005). However, the P-loop in "structural" domain 2 is highly truncated compared to that in "structural" domain 1 (Figure 5), and it explains why acetyl CoA (or CoA) is bound to "structural" domain 1 only.

In the ligand-bound structure of *M. tuberculosis* Eis, acetyl CoA makes hydrophobic interactions with Phe30, Gly130, and Ile131 in "structural" domain 1 (Figure 8). The adenine ring makes a hydrogen bond to the side chain of Glu128, while the 3'-phosphoryl group interacts with the side chain of Arg134. The pyrophosphate moiety of acetyl CoA makes hydrogen bonds to the side chain of Arg99 and the backbone nitrogen atoms of Gly101, Leu103, and Arg104. The β-mercaptoethylamine unit makes hydrogen bonds with the backbone nitrogen of Val93 and the side chain of Ser127. The acetyl carbonyl oxygen atom is hydrogen-bonded to the backbone nitrogen atom of Phe90, positioning it for a nucleophilic attack by the substrate amine group. The interactions between CoA and *M. smegmatis* Eis are identical. That is, CoA makes hydrophobic interactions with Phe26, Gly124, and Ile125 in "structural" domain 1 (Figure 8). The adenine ring makes a hydrogen bond to the side chain of Glu122, while the 3'-phosphoryl group interacts with the side chain of Arg128. The pyrophosphate moiety of acetyl CoA makes hydrogen bonds to the side chain of Arg93 and the backbone nitrogen atoms of Gly95, Leu97, and Arg98. The β-mercaptoethylamine unit makes hydrogen bonds with the backbone nitrogen of Val87 and the side chain of Ser121.

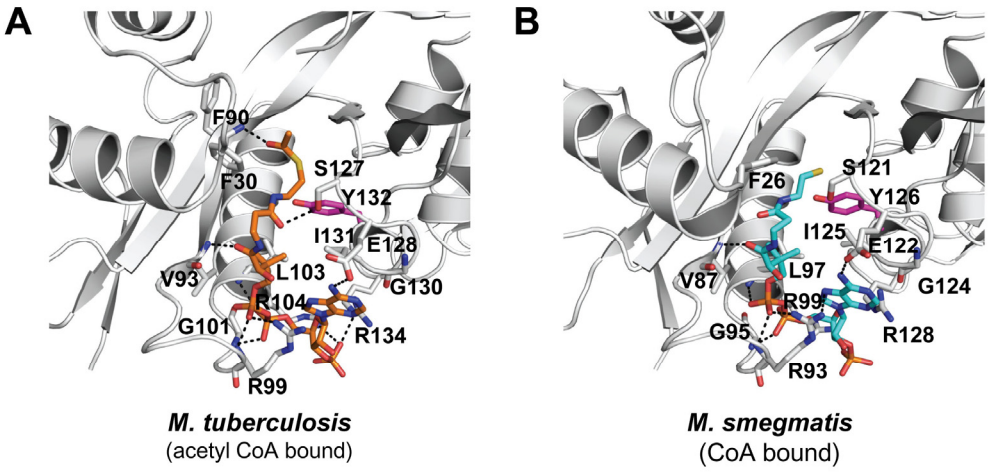


Figure 8. Structural comparison of *M. tuberculosis* Eis and *M. smegmatis* Eis

(A) Cartoon representation of ligand binding pocket in *M. tuberculosis* Eis. (B) Cartoon representation of ligand binding pocket in *M. smegmatis* Eis.. Amino acid residues around acetyl CoA or CoA are shown in a stick model. Nitrogen and oxygen atoms are colored in blue and red, respectively. Black dotted lines denote hydrogen bonds.

3.4. Comparison of aminoglycosides acetyltransferase activity of Eis Proteins from *M. tuberculosis* and *M. smegmatis*

Recently, *M. tuberculosis* Eis was shown to acetylate multiple amines of many aminoglycosides, including the second line-injectable antituberculosis drugs kanamycin and amikacin (Zaunbrecher et al., 2009; Chen et al., 2011). It was also argued that kanamycin may not be a natural substrate of *M. tuberculosis* Eis because of a high K_m value with it (Zaunbrecher et al., 2009). To examine whether there is a functional difference between Eis proteins from *M. tuberculosis* and *M. smegmatis* in aminoglycoside acetylation, I have compared acetylation activities of these two Eis proteins against amikacin, kanamycin A, tobramycin, and paromomycin. Interestingly, *M. smegmatis* Eis acetylated the tested aminoglycosides as quickly as, or more rapidly than, *M. tuberculosis* Eis (Figure 9). Steady-state kinetic parameters, as measured by K_m and k_{cat} values (Table 2), indicate that the aminoglycoside acetyltransferase activity of *M. smegmatis* Eis is comparable to or higher than that of *M. tuberculosis* Eis. This result cannot explain the enhanced intracellular survival of mycobacteria by *M. tuberculosis* Eis.

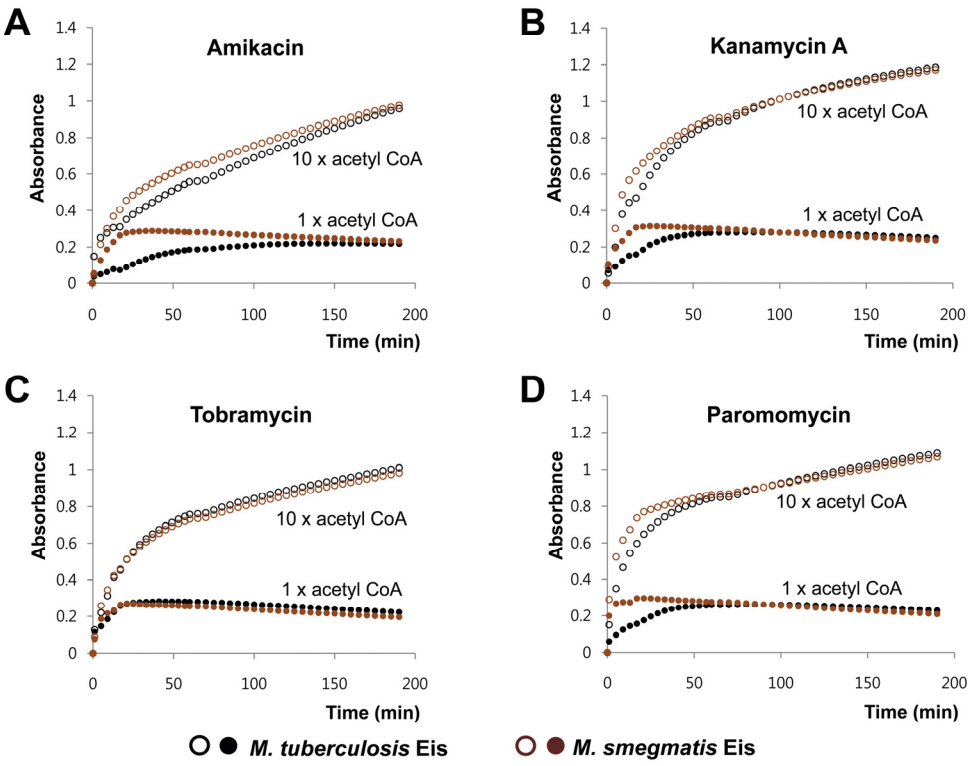


Figure 9. Acetylation assay of aminoglycosides by Eis

(A) Amikacin, (B) kanamycin A, (C) tobramycin, and (D) paromomycin. Acetylation of aminoglycosides by Eis was monitored spectrophotometrically by measuring the absorbance at 324 nm. Black open and closed circles denote the measurements by *M. tuberculosis* Eis using 10 and one equivalent(s) of acetyl CoA, respectively. Brown open and closed circles denote the measurements by *M. smegmatis* Eis using 10 and one equivalent(s) of acetyl CoA, respectively.

Table 2. Kinetic parameters of aminoglycoside *N*-acetylation by Eis proteins

Aminoglycoside	*M. tuberculosis* Eis			*M. smegmatis* Eis		
	K_m (uM)	k_{cat} (s^{-1})	k_{cat}/K_m ($M^{-1}s^{-1}$)	K_m (uM)	k_{cat} (s^{-1})	k_{cat}/K_m ($M^{-1}s^{-1}$)
Amikacin	73	0.017	233	61	0.025	409
Kanamycin A	81	0.032	395	52	0.041	788
Tobramycin	71	0.179	2,521	69	0.182	2,638
Paramomycin	85	0.032	376	48	0.063	1,313

3.5. Comparison of overall structure of Eis proteins from *M. tuberculosis* and *M. smegmatis*

To understand the observed catalytic properties, I have determined and compared the crystal structures of both *M. tuberculosis* and *M. smegmatis* Eis proteins. The crystal structure of selenomethionine-substituted *M. tuberculosis* Eis in the acetyl CoA-bound form was determined by *de novo* phasing using the single anomalous diffraction data to 2.80 Å. This model was used to solve the structures of *M. tuberculosis* Eis in the apo form at 2.46 Å and *M. smegmatis* Eis in the CoA-bound form at 2.03 Å by molecular replacement. The overall monomeric and hexameric structures of *M. tuberculosis* and *M. smegmatis* Eis proteins are similar to each other (Figure 4, Figure 7). That is, each monomer of both Eis proteins comprises three "structural" domains and six subunits are associated to form a hexamer of 32 symmetry. "Structural" domain 1 adopts the GNAT fold, as predicted. Unexpectedly, "structural" domain 2 is also folded into the GNAT structure despite an apparent lack of sequence similarity to other GNAT enzymes, including the Eis domain 1. "Structural" domain 3 resembles sterol carrier protein-2 (SCP-2). Either acetyl coenzyme A (CoA) or CoA is observed to be bound to "structural" domain 1 only, but not to "structural" domain 2, in the ligand-bound structures of both Eis proteins (Figure 4). The active sites

of *M. tuberculosis* and *M. smegmatis* Eis proteins are large and deep enough to accommodate aminoglycosides. However, they display distinct structural features that may explain the observed functional difference against peptide substrates.

3.6. Identification of DUSP16/MKP-7 as the N^{ε}-acetylation target of *M. tuberculosis* Eis

Besides aminoglycosides, *M. tuberculosis* Eis was previously shown to acetylate free histone proteins, but not the histone proteins in a nucleosomal complex (Crossman, 2007). None of these acetylation targets can explain the enhanced intracellular survival of mycobacteria by *M. tuberculosis* Eis; *M. tuberculosis* Eis likely has other unidentified protein targets for acetylation. Therefore, I was interested in identifying physiologically more important protein acetylation targets of *M. tuberculosis* Eis. *M. tuberculosis* Eis enhances macrophage survival through the regulation of JNK-dependent ROS signaling (Shin et al., 2010), while DUSP16/MKP-7 works as a JNK-specific phosphatase *in vivo* (Masuda et al., 2001). Acetylation at Lys57 of DUSP1/MKP-1, a nuclear-localized phosphatase that inactivates MAPK members by dephosphorylation, promoted the interaction of DUSP1/MKP-1

with its substrate p38 MAPK and inhibited innate immune signaling (Cao et al., 2008; Chi and Flavell, 2008). Three major subfamilies of MAPKs are ERKs, p38 MAPKs, and JNKs. These MAPKs are activated by MAPK kinases, which are in turn activated by a set of MAPK kinase kinases. The MAPK pathways that mediate innate immune signaling include MAPK kinases 3/4/6, p38, and JNKs (Hegde et al., 2001; and references therein).

On the basis of these reports, I spectulated that *M. tuberculosis* Eis may acetylate component(s) of the MAPK signaling pathways, such as MAPK phosphatases, to negatively control autophagy, phagosome maturation, and ROS generation, ultimately leading to the suppression of host immune responses. Therefore, I examined whether *M. tuberculosis* Eis may acetylate DUSP16/MKP-7, which is known to be a JNK-specific phosphatase (Masuda et al., 2001). I tested a peptide within the MAPK-docking domain of DUSP16/MKP-7: ^{53}LMKRRLQQDKVLIT^{66} [MKP-7(53–66)] (Cao et al., 2008). The underlined Lys62 was predicted to be the acetylation site in the DUSP16/MKP-7 docking domain (Cao et al., 2008). I also tested a peptide within the MAPK-docking domain of DUSP1/MKP-1: ^{50}TIVRRRAKGAMGLE^{63} [MKP-1(50–63)] (Cao et al., 2008). The underlined lysine was established as the acetylation site in the DUSP1/MKP-

1 docking domain (Cao et al., 2008). With these peptides as possible substrates, I performed *in vitro* acetylation assays using either *M. tuberculosis* Eis or *M. smegmatis* Eis.

The MALDI-TOF mass spectra of the MKP-7(53–66) peptide show that the unmodified peptide peak at 1,741 Da is shifted to 1,783 Da after *in vitro* acetylation for 5 min by *M. tuberculosis* Eis (Figure 10). The observed increase of 42 Da in the peptide mass corresponds to covalent attachment of an acetyl group (CH_3CO–). On the other hand, the mass (1,557 Da) of the MKP-1(50–63) peptide before and after *in vitro* acetylation for 30 min by *M. tuberculosis* Eis is identical to the predicted mass of the unmodified peptide (Figure 10). These results indicate that *M. tuberculosis* Eis quickly attaches a single acetyl group to the MKP-7(53–66) peptide, whereas it does not readily acetylate the MKP-1(50–63) peptide. The MKP-7(53–66) peptide has three potential *N*-acetylation sites: one N^{α}-acetylation site at the amino terminus and two N^{ε}-acetylation sites at Lys55 and Lys62. In order to identify which of the three possible acetylation sites was actually modified, I performed *de novo* sequencing of the acetylated MKP-7(53–66) peptide by MALDI tandem mass spectrometry (Figure 10). Peaks of unacetylated a1, a2, y10, and y11 ions are present at 86.009, 217.157, 1214.653, and 1369.811

Da, respectively, while peaks of acetylated b-ion series b3*, b4*, b5*, and b8* are present at 415.398, 571.565, 727.680, and 1096.718 Da, respectively. This result identifies Lys55 of the MKP-7(53–66) peptide as the N^{ε}-acetylation site. By analogy with DUSP1/MKP-1 (Cao et al., 2008), I suggest that N^{ε}-acetylation of DUSP16/MKP-7 at Lys55 by *M. tuberculosis* Eis may increase the interactions between DUSP16/MKP-7 and JNK by neutralizing the positive charge within the docking domain of DUSP16/MKP-7. *M. smegmatis* Eis acetylated both MKP-1(50–63) and MKP-7(53–66) peptides very quickly in about 5 min, resulting in a mass increase by 42 Da. (Figure 10). I have confirmed by MALDI tandem mass spectrometry that these peptides are N^{α}-acetylated at the amino terminus by *M. smegmatis* Eis (Figure 11). My data thus establish that Eis proteins from both *M. tuberculosis* and *M. smegmatis* are catalytically active as aminoglycoside *N*-acetyltransferases but only *M. tuberculosis* Eis acts as an efficient N^{ε}-acetyltransferase that can acetylate DUSP16/MKP-7.

I further tested the protein *N*-acetyltransferase activity using the recombinant human DUSP16/MKP-7. I could express and purify DUSP/MKP-7(1–153), which encompasses the docking domain, whereas the full-length DUSP16/MKP-7 was not expressed in *E. coli*. When

DUSP16/MKP-7(1−153) was incubated with the wild-type *M. tuberculosis* Eis in the presence of [^{14}C]-acetyl CoA, it was efficiently acetylated by *M. tuberculosis* Eis (Figure 12). However, acetylation of DUSP16/MKP-7(1−153) by *M. smegmatis* Eis was much less efficient (Figure 12). I next performed mutation experiments to confirm the acetylation site(s) in DUSP16/MKP-7(1−153). Each of Lys52, Lys55, and Lys62 in DUSP16/MKP-7(1−153) was replaced by arginine and the acetylation reaction was carried out. Mutation of Lys55 drastically reduced the level of acetylation by *M. tuberculosis* Eis, whereas mutations of Lys52 and Lys62 showed no such drastic reduction (Figure 12). With *M. smegmatis* Eis, mutation of Lys55 had little effect on the level of acetylation, suggesting that acetylation by *M. smegmatis* Eis likely occurred at the amino terminus, but not at Lys55 (Figure 12). These results strongly suggest that *M. tuberculosis* Eis functions as a protein N^{ε}-acetyltransferase toward human DUSP/MKP-7.

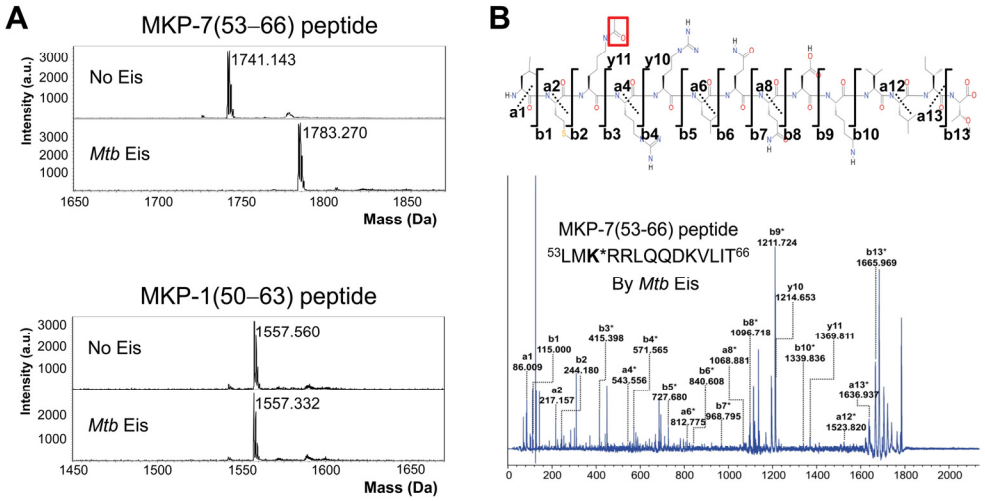


Figure 10. Mass analyses of MKP-7(53–66) and MKP-1(50–63) peptides

(A) Mass spectra of MKP-7(53–66) and MKP-1(50–63) peptides before (top panel) and after acetylation by *M. tuberculosis* Eis (middle) and *M. smegmatis* Eis (bottom). (B) MALDI MS/MS spectrum of the MKP-7(53–66) peptide acetylated by *M. tuberculosis* Eis. The fragments marked with an asterisk (*) are +42 Da-shifted ions, as compared to the counterparts that would be generated from the unmodified peptide. The acetylated MKP-7(53–66) fragmentation notation using the scheme of Roepstorff and Fohlman is given above the spectrum (Roepstorff and Fohlman, 1984). The acetyl group of modified Lys55 is highlighted by enclosing in a red box.

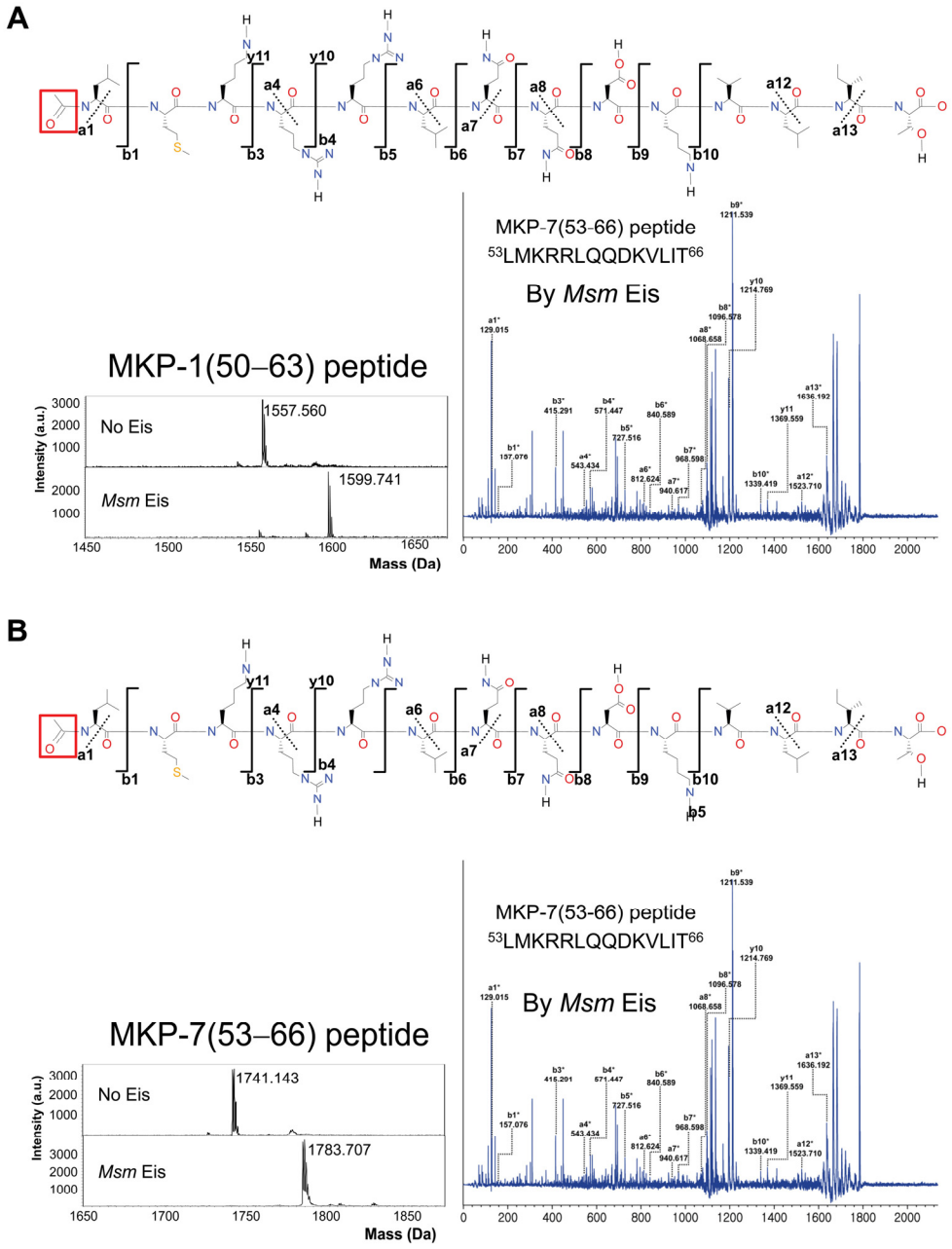


Figure 11. MALDI MS/MS spectra of the peptides acetylated by *M. smegmatis* Eis

(A) MKP-1(50−63) peptide and (B) MKP-7(53−66) peptide. The fragments marked with an asterisk (*) are +42 Da-shifted ions, as compared to the counterparts that would be generated from the unmodified peptide. The fragmentation notation using the scheme of Roepstorff and Fohlman is given above each spectrum (Roepstorff and Fohlman, 1984). The acetyl group attached to the N-terminus is enclosed in a red box.

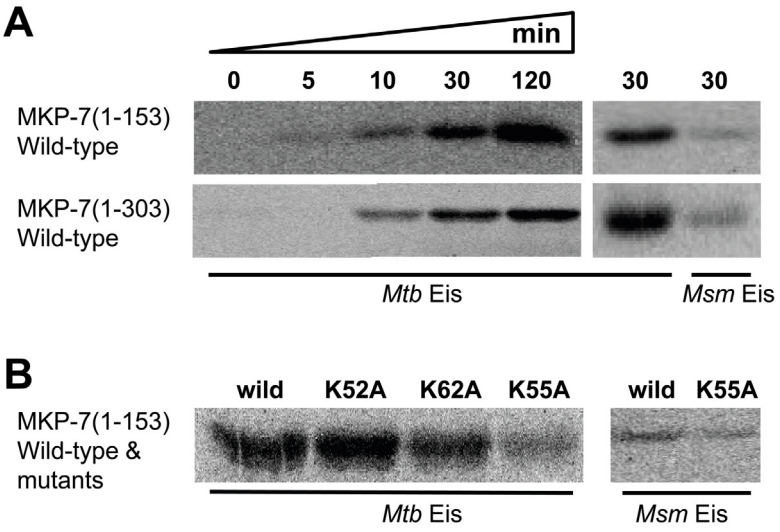


Figure 12. Acetyltransferase activity assay of *M. tuberculosis* Eis and *M. smegmatis* Eis using the wild-type and mutants of recombinant human DUSP/MKP-7(1–153) as potential substrates

(A) Time-course acetyltransferase activity assay of *M. tuberculosis* Eis and a comparison of *M. tuberculosis* Eis and *M. smegmatis* Eis activities. The wild-type DUSP16/MKP-7(1–153) was incubated with [^{14}C]-labeled acetyl CoA and Eis for the indicated duration at 37°C. The reaction products were separated by 15%(w/v) SDS-PAGE and the acetylated protein bands were visualized using a Bioimage analyzer. (B) Acetyltransferase activity of *M. tuberculosis* Eis and *M. smegmatis* Eis toward the wild-type and mutants of DUSP16/MKP-7(1–153). All reactions were carried out at 37°C for 30 min.

3.7. Comparison of substrate binding sites in Eis proteins

Biochemical studies indicated that both *M. tuberculosis* and *M. smegmatis* Eis proteins could readily acetylate aminoglycosides, with *M. smegmatis* Eis being a marginally better catalyst as an aminoglycoside *N*-acetyltransferase (Figure 9). In contrast, their substrate preferences against MKP-7(53–66) and MKP-1(50–63) peptides were different. *M. tuberculosis* Eis acetylated an internal lysine (Lys55) of the MKP-7(53–66) peptide rapidly, whereas *M. smegmatis* Eis preferentially acetylated the terminal amino group of these peptides (Figure 10, Figure 11). To understand the observed difference in peptide acetylation site preferences, I have compared the active site features of these Eis proteins. Both Eis active sites show negative electrostatic potential surfaces (Figure 13) and key residues around the thiol group of acetyl CoA (or CoA) are identical (Asp32/Tyr132/Phe408 in *M. tuberculosis*; Asp28/Tyr126/Phe402 in *M. smegmatis*). Interestingly, I noticed a difference in the shape of the predicted substrate binding sites adjacent to the bound ligand in their active sites. *M. tuberculosis* Eis has a deep and narrow channel, whereas *M. smegmatis* Eis has a deep and round pocket (Figure 13). This structural difference in the substrate binding sites is mainly due to different arrangements of helix α2 (Gly35–Leu45 in *M.*

tuberculosis Eis; Glu32–Met41 in *M. smegmatis*), part of "structural" domain 1, and helix α6 (Gln200–Cys210 in *M. tuberculosis* Eis; Asp190–Ala198 in *M. smegmatis*) and the following loop, which are part of "structural" domain 2 (Figure 14). The narrow channel of *M. tuberculosis* Eis is formed by limited hydrophobic interactions between these helices. In comparison, the round pocket-shaped substrate binding site of *M. smegmatis* Eis is formed by extensive interactions between helix α2 in "structural" domain 1 and helix α6 followed by an additional 3_{10} helix α6' in "structural" domain 2 (Figure 4). The 3_{10} helix α6' is present in *M. smegmatis* Eis only (Figure 4). The close contact between α2 and α6 helices in *M. smegmatis* Eis is stabilized by hydrogen bondings of Gln33-Arg200 and Thr34-Asp195 pairs (Figure 14). In addition, the side chain of Trp42 in *M. tuberculosis* Eis is located between the side chains of Trp19 (on α1) and Phe90 (on β4) in "structural" domain 1, while the corresponding residue in *M. smegmatis* Eis, Trp38, is located on the other side of Phe84 and makes van der Waals contacts with the side chains of Met41 (on α2), Leu192 (on α6), and Tyr400 (Figure 14). The elongated substrate-binding channel in *M. tuberculosis* Eis appears to be suitable not only for accommodating aminoglycosides but also for recognizing the polypeptide substrate in a sequence-specific manner. The deep, round-shaped substrate-binding pocket in *M. smegmatis* Eis appears

more suitable for accommodating aminoglycosides and the terminal amino group of peptides than sequence-specific recognition of polypeptides.

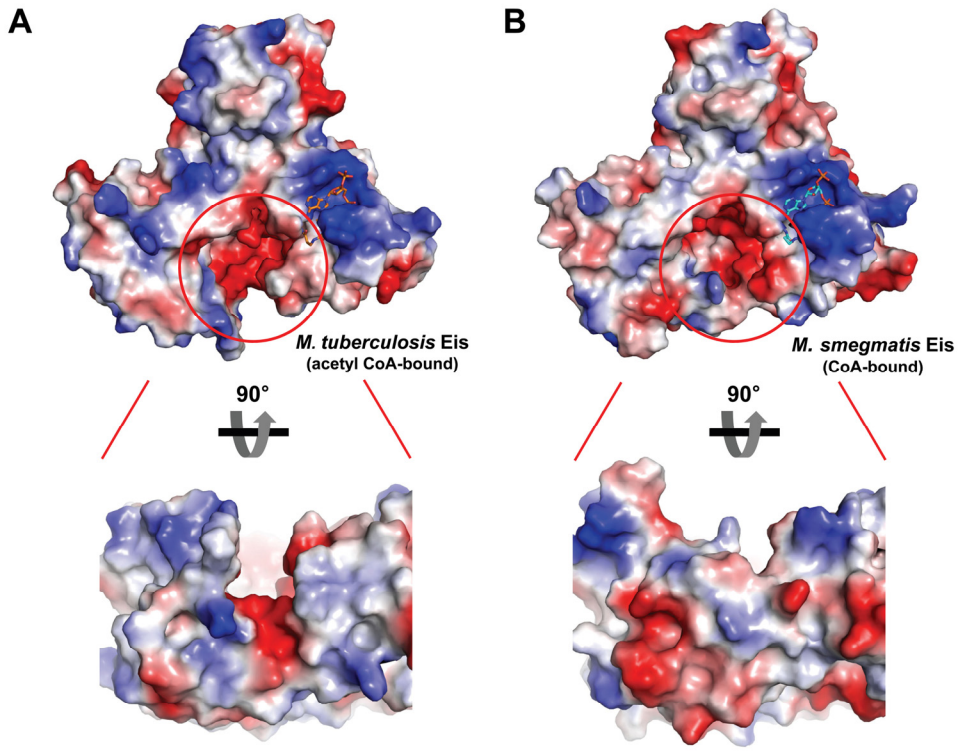


Figure 13. Structures of *M. tuberculosis* and *M. smegmatis* Eis

(A) Electrostatic potential at the surface of *M. tuberculosis* Eis monomer, with an enlarged view from an orthogonal angle. (B) Electrostatic potential at the surface of *M. smegmatis* Eis monomer, with an enlarged view from an orthogonal angle. Blue and red colors correspond to positive and negative potentials, respectively. Red circles indicate possible substrate binding sites of *M. tuberculosis* and *M. smegmatis* Eis.

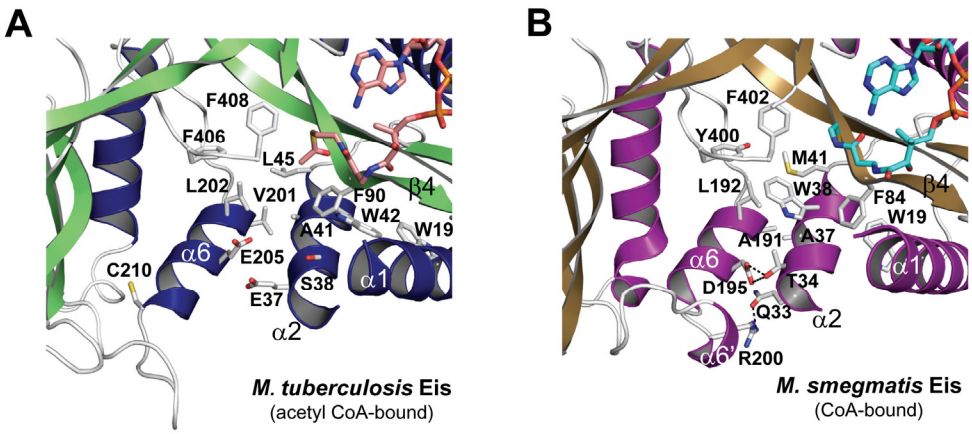


Figure 14. Structural comparison of substrate binding site in *M. tuberculosis* Eis and *M. smegmatis* Eis

(A) Residues around the substrate binding sites of *M. tuberculosis* Eis (B) Residues around the substrate binding sites of *M. smegmatis* Eis. These close-up views roughly correspond to the regions indicated by red circles in Figure 13. Side chains of amino acid residues around the substrate binding sites are shown in a stick model. Nitrogen and oxygen atoms are colored in blue and red, respectively. Black dotted lines denote hydrogen bonds. All figures are drawn in the same view.

3.8. *M. tuberculosis* Eis, but not *M. smegmatis* Eis, inhibits lipopolysaccharide-induced JNK phosphorylation.

As DUSP16/MKP-7 can dephosphorylate JNK in macrophages in response to lipopolysaccharide (LPS) (Tanoue et al., 2001; Han et al., 2002; Matsuguchi et al., 2001; Weston and Davis, 2007; Kondoh and Nishida, 2007), I examined whether *M. tuberculosis* Eis or *M. smegmatis* Eis inhibits LPS-induced JNK activation in bone marrow-derived macrophages (BMDMs). *M. tuberculosis* Eis significantly reduced the level of phosphorylated JNK in BMDMs in response to LPS, whereas *M. smegmatis* Eis did not (Figure 15). These data, together with previous finding that *M. tuberculosis* Eis acetylates DUSP16/MKP-7 at Lys55, indicate that *M. tuberculosis* Eis suppresses host innate immune responses through inactivation of JNK via acetylation of DUSP16/MKP-7.

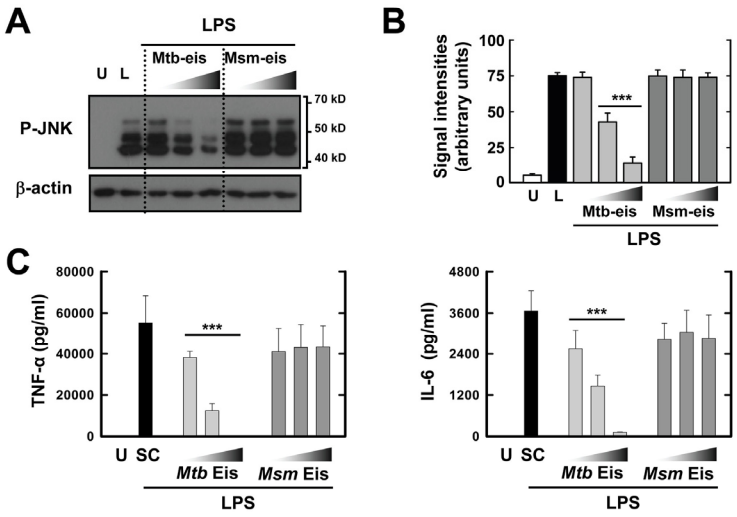


Figure 15. *M. tuberculosis* Eis, but not *M. smegmatis* Eis, significantly inhibits JNK activation in BMDM cells

Cells were treated with or without *M. tuberculosis* Eis (or *M. smegmatis* Eis) (5, 10, or 20 ug ml^{-1}) for 1 h, followed by stimulation with LPS (100 ng ml^{-1}) for 30 min. Cells were then harvested, lysed, and subjected to Western blot analysis using antibodies raised to phospho-JNK and β-actin. Data shown are representative of three independent experiments that all yielded similar results. (A) Expression of phospho-JNK and β-actin in cytoplasmic extracts of BMDMs was quantified densitometrically (B). Data represent the mean ± standard deviation and are representative of three independent experiments. ***$p < 0.001$, vs. LPS-stimulated condition; U, untreated; L, LPS-treated condition without *M. tuberculosis* Eis or *M. smegmatis* Eis.

3.9. Mechanism of acetylation by *M. tuberculosis* Eis

The ternary *Tetrahymena* Gcn5 (tGcn5)/CoA/histone H3 complex allows one to derive a detailed mechanism of catalysis that is mediated by the glutamate within the core domain of Gcn5 (Rojas et al., 2009). Glu407 of *M. tuberculosis* Eis is bound to a water molecule located between the Glu407 side chain and the reactive Lys55 residue of the DUSP16/MKP-7. This water molecule is ideally located to shuttle a proton from the reactive Lys55 of the histone to Glu407 of the protein. The lysine amino group to be acetylated is ideally positioned by the His125 backbone for a nucleophilic attack on the CoA thioester. The main chain amide nitrogens of Phe 90 and Val 91 donate a hydrogen bond to the carbonyl oxygen of the thioester in this model, thus polarizing the thioester carbonyl and statbilizing the tetrahedral transtion state (Chen et al., 2011). The C-terminal carboxylate of Phe408 interacts with the amino group through a bridging water molecule and likely serves as a remote base. The conserved Tyr132 at a distance of 3.6 Å from the sulfhydryl group of CoA likely serves as a general acid to protonate the CoA thiolate (Figure 16) (Chen et al., 2011).

3.10. Mechanism of host immune suppression by *M. tuberculosis* Eis

Survival of *M. tuberculosis* inside human macrophages is central to tuberculosis infection, latency, disease activation, and transmission (Russell,

2001). The most important survival strategies of the pathogen are the inhibition of phagosomal maturation and autophagy (Gutierrez et al., 2004). Autophagy has been recognized as innate and adaptive immune defense mechanisms (Gutierrez et al., 2004). A functional overlap was suggested between phagosome maturation and autophagy, both of which depend on Beclin 1 and VPS34 as key players (Gutierrez et al., 2004). Beclin 1, a pro-autophagy BH3 domain-containing protein, plays a central role in autophagosome formation by interacting with several other factors to promote the formation of Beclin 1-VPS34-VPS15 core complexes (Kang et al., 2011; Funderburk et al., 2010). In mammalian cells under nonstarvation conditions, the anti-apoptotic protein Bcl-2 binds to Beclin 1 and inhibits its autophagy function (Wei et al., 2008). Starvation induces Bcl-2 dissociation from Beclin 1, via phosphorylation of Bcl-2, and autophagy activation. Furthermore, JNK1, but not JNK2, was found to mediate starvation-induced Bcl-2 phosphorylation (Wei et al., 2008). Recently, JNK signaling was shown to mediate amplification of ROS production during multiple stresses including infection (Chambers and LoGrasso, 2011). Cellular stress altered mitochondria, causing JNK to translocate to the mitochondria and to amplify up to 80% of the ROS generated largely by Complex I. ROS activates JNK via a sequence of events for JNK mitochondrial signaling (Chambers and

LoGrasso, 2011). There exists a molecular cross talk between autophagy and apoptosis, because Beclin 1 can be cleaved by caspases and its pro-autophagic activity is lost (Djavaheri-Mergny et al., 2010). Moreover, the resulting C-terminal fragment of Beclin 1 could amplify mitochondrion-mediated apoptosis (Djavaheri-Mergny et al., 2010; Wirawan et al., 2010).

The kinase activity of MAPKs such as JNK is negatively regulated by DUSPs (also called MKPs). DUSP16/MKP-7 was found to work as a JNK-specific phosphatase *in vivo*, because forced expression of DUSP16/MKP-7 suppressed activation of MAPKs in the order of selectivity, JNK >> p38 > ERK (Masuda et al., 2001). When expressed in mammalian cells, DUSP16/MKP-7 was localized exclusively in the cytoplasm (Masuda et al., 2001). Acetylation of the components of MAPK pathways on serine/threonine and lysine residues was previously reported to serve as a regulatory mechanism in biological signaling. *Yersinia* YopJ was shown to act as an acetyltransferase to modify serine and threonine residues in the activation loop of MAPK kinase-6 and thereby blocking phosphorylation and subsequent activation of the kinase activity (Mukherjee et al., 2006; Mukherjee et al., 2007). A nuclear-localized DUSP1/MKP-1 is acetylated by p300 on Lys57 within its docking domain, resulting in deactivation of Toll-like receptor inflammatory signaling and inhibition of innate immune

signaling (Cao et al., 2008). p300 contains a histone acetyltransferase domain. Acetylation of DUSP1/MKP-1 enhanced its interaction with p38, thereby increased its phosphatase activity, and interrupted MAPK signaling cascade (Cao et al., 2008; Chi and Flavell, 2008). In this study, I have discovered that *M. tuberculosis* Eis acts as an N^{ε}-acetyltransferase to acetylate Lys55 within the docking domain of DUSP16/MKP-7. I have also shown that *M. tuberculosis* Eis, but not *M. smegmatis* Eis, significantly down-regulated the LPS-induced JNK phosphorylation. On the basis of these findings, I propose that acetylation of DUSP16/MKP-7 by *M. tuberculosis* Eis is the key initial event in the JNK-dependent inhibition of autophagy, phagosome maturation, and ROS generation, which ultimately contributes to enhanced survival of *M. tuberculosis* within the macrophage cells.

It has been well established that protein lysine acetylation critically regulates gene transcription by targeting histones as well as a variety of transcription factors in the nucleus (Close et al., 2010). Numerous proteins located outside the nucleus have also been demonstrated to be acetylated (Close et al., 2010). Indeed, protein lysine acetylation is emerging as a major mechanism by which key proteins are regulated in many physiological processes (Close et al., 2010). Recent reports also link lysine acetylation to

heterochromatin assembly, sister chromatid cohesion, cytoskeleton dynamics, autophagy, receptor signaling, RNA processing, and metabolic control (Kim and Yang, 2011). Proteomics studies indicate that the complexity of the acetylome potentially rivals that of the phosphoproteome (Norris et al., 2009). Therefore, it is not surprising to find that *M. tuberculosis* utilizes Eis to acetylate the host signaling protein DUSP16/MKP-7 in suppressing immune responses for its survival in macrophages. *M. tuberculosis* might have evolved in such as way that its *eis* gene product has retained much of the aminoglycoside *N*-acetyltransferase activity, while it has gained a significantly higher protein lysine N^{ϵ}-acetyltransferase activity to disrupt the cellular signaling pathway for intracellular survival.

Similarly to *M. tuberculosis* Eis, protein kinase G (PknG) of *M. tuberculosis* inhibits phagosome-lysosome fusion and mediates intracellular survival of mycobacteria by disrupting the host cellular signaling (Walburger et al., 2004). PknG is one of the 11 eukaryotic-like Ser/Thr protein kinases encoded by the *M. tuberculosis* genome and is secreted within macrophages. *M. tuberculosis* Eis, like *M. tuberculosis* PknG, could be an excellent target for the development of drugs that induce mycobacterial death inside macrophages. An advantage of targeting Eis or PknG is that it does not kill

the bacteria per se but instead facilitates the macrophage to carry out its natural antibacterial activity, delivering intracellularly surviving mycobacteria to lysosomes for destruction (Nguyen and Pieters, 2009). Another potential advantage of targeting a secreted protein such as Eis or PknG is that its inhibitors are not required to be transported through the extremely impermeable mycobacterial cell wall. This may greatly improve the bactericidal activity of the compounds (Nguyen and Pieters, 2009). *M. tuberculosis* Eis has an extra advantage, because existing aminoglycoside drugs such as kanamycin and amikacin can be made effective by inhibiting its aminoglycoside *N*-acetyltransferase activity. Structural information would be useful in structure-based discovery of peptidomimetic or small molecule inhibitors that target the active site of *M. tuberculosis* Eis.

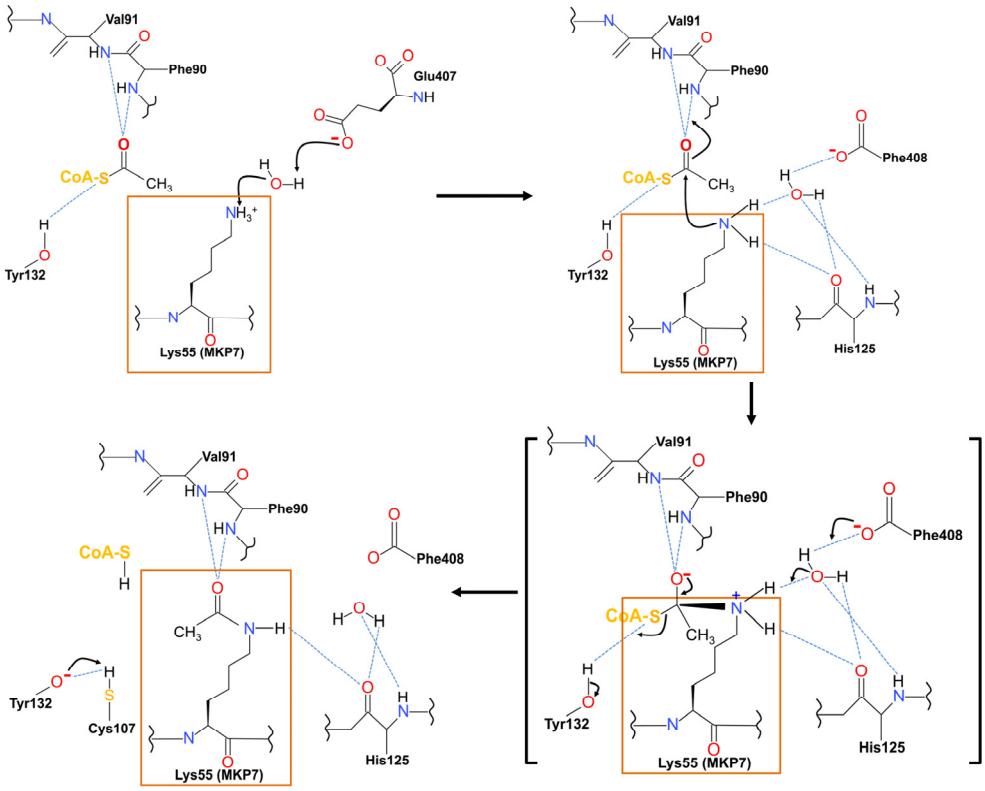


Figure 16. Proposed reaction mechanism for *M. tuberculosis* Eis mediated catalysis of DUSP16/MKP-7

4. References

Adams P. D., Afonine P. V., Bunkóczi G., Chen V. B., Davis I. W., Echols N., Headd J. J., Hung L. W., Kapral G. J., Grosse-Kunstleve R. W., McCoy A. J., Moriarty N. W., Oeffner R., Read R. J., Richardson D. C., Richardson J. S., Terwilliger T. C., Zwart P. H. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr Sect D Biol Crystallogr* 66: 213–221.

Brünger A. T., Adams P. D., Clore G. M., DeLano W. L., Gros P., Grosse-Kunstleve R. W., Jiang J. S., Kuszewski J., Nilges M., Pannu N. S., Read R. J., Rice L. M., Simonson T., Warren G. L. (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr Sect D Biol Crystallogr* 54: 905–921.

Cao W., Bao C., Padalko E., Lowenstein C. J. (2008) Acetylation of mitogen-activated protein kinase phosphatase-1 inhibits Toll-like receptor signaling. *J Exp Med* 205: 1491–1503.

Chambers J. W., LoGrasso P. V. (2011) Mitochondrial c-Jun N-terminal kinase (JNK) signaling initiates physiological changes resulting in

amplification of reactive oxygen species generation. *J Biol Chem* 286: 16052–16062.

Chen V. B., Arendall W. B. 3rd, Headd J. J., Keedy D. A., Immormino R. M., Kapral G. J., Murray L. W., Richardson J. S., Richardson D. C. (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* 66: 12–21.

Chen W., Biswas T., Porter V. R., Tsodikov O. V., Garneau-Tsodikova S. (2011) Unusual regioversatility of acetyltransferase Eis, a cause of drug resistance in XDR-TB. *Proc Natl Acad Sci USA* 108: 9804–9808.

Chi H., Flavell R. A. (2008) Acetylation of MKP-1 and the control of inflammation. *Science Signal* 1: pe44.

Chiang C. Y., Centis R., Migliori G. B. (2010) Drug-resistant tuberculosis: past, present, future. *Respirology* 15: 413–432.

Clements A, Rojas J. R., Trievel R. C., Wang L., Berger S. L., Marmorstein R. (1999) Crystal structure of the histone acetyltransferase domain of the human PCAF transcriptional regulator bound to coenzyme A. EMBO J 18: 3521–3532.

Close P., Creppe C., Gillard M., Ladang A., Chapelle J. P., Nguyen L., Chariot A. (2010) The emerging role of lysine acetylation of non-nuclear proteins. *Cell Mol Life Sci* 67: 1255–1264.

Collaborative Computational Project, Number 4 (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr Sect D Biol Crystallogr* 50: 760–763.

Crossman D. K. (2007) Characterization of a novel acetylatransferase found only in pathogenic strains of *Mycobacterium tuberculosis*. Ph.D. thesis. The University of Alabama, USA.

Dahl J. L., Wei J., Moulder J. W., Laal S., Friedman R. L. (2001) Subcellular localization of the intracellular survival-enhancing Eis protein of *Mycobacterium tuberculosis*. *Infect Immun* 69: 4295–4302.

Djavaheri-Mergny M., Maiuri M. C., Kroemer G. (2010) Cross talk between apoptosis and autophagy by caspase-mediated cleavage of Beclin 1. *Oncogene* 29: 1717–1719.

Dundas J., Ouyang Z., Tseng J., Binkowski A., Turpaz Y., Liang J. (2006) CASTp: computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues. *Nucleic Acids Res* 34: W116–W118.

Dye C., Williams B. G. (2010) The population dynamics and control of tuberculosis. *Science* 328: 856–861.

Dyer D. H., Lovell S, Thoden J. B., Holden H. M., Rayment I., Lan Q. (2003) The structural determination of an insect sterol carrier protein-2 with a ligand-bound C16 fatty acid at 1.35-Å resolution. *J Biol Chem* 278: 39085–39091.

Ehrt S., Schnappinger D. (2009) Mycobacterial survival strategies in the phagosome: defence against host stresses. *Cell Microbiol* 11: 1170–1178.

Emsley P., Cowtan K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr Sect D Biol Crystallogr* 60: 2126–2132.

Funderburk S. F., Wang Q. J., Yue Z. (2010) The Beclin 1–VPS34 complex – at the crossroads of autophagy and beyond. *Trends Cell Biol* 20: 355–362.

García F. L., Szyperski T., Dyer J. H., Choinowski T., Seedorf U., Hauser H., Wüthrich K. (2000) NMR structure of the sterol carrier protein-2: implications for the biological role. *J Mol Biol* 295: 595–603.

Getahun H., Gunneberg C., Granich R., Nunn P. (2010) HIV infection-associated tuberculosis: the epidemiology and the response. *Clin Infect Dis* 50: S201–S207.

Goroncy A. K., Murayama K., Shirouzu M., Kuramitsu S., Kigawa T., Yokoyama S. (2010) NMR and X-ray structures of the putative sterol carrier protein 2 from *Thermus thermophilus* HB8 show conformational changes. *J Struct Funct Genomics* 11: 247–256.

Groeger G., Quiney C., Cotter T. G. (2009) Hydrogen peroxide as a cell-survival signaling molecule. *Antioxid Redox Signal* 11: 2655–2671.

Gutierrez M. G., Master S. S., Singh S. B., Taylor G. A., Colombo M. I., Deretic V. (2004) Autophagy is a defense mechanism inhibiting BCG and *Mycobacterium tuberculosis* survival in infected macrophages. *Cell* 119:753–766.

Haapalainen A. M,. van Aalten D. M., Meriläinen G., Jalonen J. E., Pirilä P., Wierenga R. K., Hiltunen J. K., Glumoff T. (2001) Crystal structure of the liganded SCP-2-like domain of human peroxisomal multifunctional enzyme type 2 at 1.75 Å resolution. *J Mol Biol* 313: 1127–1138.

Han S. Y., Kim S. H., Heasley L. E. (2002) Differential gene regulation by specific gain-of-function JNK1 proteins expressed in Swiss 3T3

fibroblasts. *J Biol Chem* 277: 47167–47174.

Hegde S. S., Javid-Majd F., Blanchard J. S. (2001) Overexpression and mechanistic analysis of chromosomally encoded aminoglycoside 2'-N-acetyltransferase (AAC(2')-Ic) from *Mycobacterium tuberculosis*. *J Biol Chem* 276: 45876–45881.

Holm L., Rosenström P. (2010) Dali server: conservation mapping in 3D. *Nucleic Acids Res* 38: W545–W549.

Jo E. K., Yang C. S., Choi C. H., Harding C. V. (2007) Intracellular signalling cascades regulating innate immune responses to *Mycobacteria*: branching out from Toll-like receptors. *Cell Microbiol* 9: 1087–1098.

Kang R., Zeh H. J., Lotze M. T., Tang D. (2011) The Beclin 1 network regulates autophagy and apoptosis. *Cell Death Differ* 18: 571–580.

Kim G. W., Yang X. J. (2011) Comprehensive lysine acetylomes emerging from bacteria to humans. *Trends Biochem Sci* 36: 211–220.

Kondoh K., Nishida, E. (2007) Regulation of MAP kinases by MAP kinase phosphatases. *Biochim Biophys Acta* 1773: 1227–1237.

Lazarow P. B. (2006) The import receptor Pex7p and the PTS2 targeting sequence. *Biochim Biophys Acta* 1763: 1599–1604.

Lella R. K., Sharma C. (2007) Eis (enhanced intracellular survival) protein of *Mycobacterium tuberculosis* disturbs the cross regulation of T-cells. *J Biol Chem* 282: 18671–18675.

Lensink M. F., Haapalainen A. M., Hiltunen J. K., Glumoff T., Juffer A. H. (2002) Response of SCP-2L domain of human MFE-2 to ligand removal: binding site closure and burial of peroxisomal targeting signal. *J Mol Biol* 323: 99–113.

Lin P. L., Flynn J. L. (2010) Understanding latent tuberculosis: a moving target. *J Immunol* 185: 15–22.

Masuda K., Shima H., Watanabe M., Kikuchi K. (2001) MKP-7, a novel mitogen-activated protein kinase phosphatase, functions as a shuttle protein. *J Biol Chem* 276: 39002–39011.

Matsuguchi T., Musikacharoen T., Johnson T. R., Kraft A. S., Yoshikai Y. (2001) A novel mitogen-activated protein kinase phosphatase is an important negative regulator of lipopolysaccharide-mediated c-Jun N-terminal kinase activation in mouse macrophage cell lines. *Mol Cell Biol* 21: 6999–7009.

McCoy A. J., Grosse-Kunstleve R. W., Adams P. D., Winn M. D., Storoni L. C., Read R. J. (2007) Phaser crystallographic software. *J Appl Cryst* 40: 658–674.

Meena L. S., Rajni (2010) Survival mechanisms of pathogenic *Mycobacterium tuberculosis* H37Rv. *FEBS J* 277: 2416–2427.

Miller R., Gallo S. M., Khalak H. G., Weeks C. M. (1994) *SnB*: crystal structure determination via *Shake-and-Bake*. *J Appl Cryst* 27: 613–621.

Mukherjee S., Keitany G., Li Y., Wang Y., Ball H. L., Goldsmith E. J., Orth K. (2006). *Yersinia* YopJ acetylates and inhibits kinase activation by blocking phosphorylation. *Science* 312: 1211−1214.

Mukherjee S., Hao Y. H., Orth K. (2007). A newly discovered post-translational modification—the acetylation of serine and threonine residues. *Trends Biochem Sci* 32: 210−216.

Neuwald A. F., Landsman D. (1997) GCN5-related histone *N*-acetyltransferases belong to a diverse superfamily that includes the yeast SPT10 protein. *Trends Biochem Sci* 22: 154–155.

Nguyen L., Pieters J. (2009) Mycobacterial subversion of chemotherapeutic reagents and host defense tactics: challenges in tuberculosis drug development. *Annu Rev Pharmacol Toxicol* 49: 427–453.

Norris K. L., Lee J. Y., Yao T. P. (2009) Acetylation goes global: the emergence of acetylation biology. *Sci Signal* 2: pe76.

Otwinowski Z., Minor W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol* 276: 307–326.

Roepstorff P., Fohlman J. (1984) Proposal for a common nomenclature for sequence ions in mass spectra of peptides. *Biomed Mass Spectrom* 11: 601.

Rojas J. R., Trievel R. C., Zhou J., Mo Y., Li X., Berger S. L., Allis C. D., Marmorstein R. (1999) Structure of the Tetrahymna GCN5 bound to coenzyme-A and a histone H3 peptide. *Nature* 401: 93-98

Russell D. G. (2001) *Mycobacterium tuberculosis*: here today, and here tomorrow. *Nat Rev Mol Cell Biol* 2: 569–577.

Samuel L. P., Song C. H., Wei J., Roberts E. A., Dahl J. L., Barry C. E. 3rd, Jo E. K., Friedman R. L. (2007) Expression, production and release of the Eis protein by *Mycobacterium tuberculosis* during infection of macrophages and its effect on cytokine secretion. *Microbiology* 153: 529–540

Scheibner K. A., De Angelis J., Burley S. K., Cole P. A. (2002) Investigation of the roles of catalytic residues in serotonin *N*-acetyltransferase. *J Biol Chem* 277: 18118–18126.

Schroeder F., Atshaves B. P., McIntosh A. L., Gallegos A. M., Storey S. M., Parr R. D., Jefferson J. R., Ball J. M., Kier A. B. (2007) Sterol carrier protein-2: new roles in regulating lipid rafts and signaling. *Biochim Biophys Acta* 1771: 700–718.

Shi D., Sagar V., Jin Z., Yu X., Caldovic L., Morizono H., Allewell N. M., Tuchman M. (2008) The crystal structure of *N*-acetyl-L-glutamate synthase from *Neisseria gonorrhoeae* provides insights into mechanisms of catalysis and regulation. *J Biol Chem* 283: 7176–7184.

Shin D. M., Jeon B. Y., Lee H. M., Jin H. S., Yuk J. M., Song C. H., Lee S. H., Lee Z. W., Cho S. N., Kim J. M., Friedman R. L., Jo E. K. (2010) *Mycobacterium tuberculosis* Eis regulates autophagy, inflammation, and cell death through redox-dependent signaling. *PLoS Pathog* 6: e1001230.

Shin M. , Kim H. J. (2011) Peptide C-terminal sequence analysis by MALDI-TOF MS utilizing EDC coupling with Br signature. *Bull Korean Chem Soc* 32: 1–4.

Singarapu K. K., Radek J. T., Tonelli M., Markley J. L., Lan Q. (2010) Differences in the structure and dynamics of the apo- and palmitate-ligated forms of *Aedes aegypti* sterol carrier protein 2 (AeSCP-2). *J Biol Chem* 285: 17046–17053.

Stolowich N. J., Petrescu A. D., Huang H., Martin G. G., Scott A. I., Schroeder F. (2002) Sterol carrier protein-2: structure reveals function. *Cell Mol Life Sci* 59: 193–212.

Tanoue T., Yamamoto T., Maeda R., Nishida E. (2001) A novel MAPK phosphatase MKP-7 acts preferentially on JNK/SAPK and p38 alpha and beta MAPKs. *J Biol Chem* 276: 26629–26639.

Terwilliger T. C., Berendzen J. (1999) Automated MAD and MIR structure solution. *Acta Crystallogr Sect D Biol Crystallogr* 55: 849–861.

Terwilliger T. C. (2000) Maximum-likelihood density modification. *Acta Crystallogr Sect D Biol Crystallogr* 66: 213–221.

Thompson J. D., Gibson T. J., Plewniak F., Jeanmougin F., Higgins D. G. (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* 25: 4876–4882.

Van Duyne G. D., Standaert R. F., Karplus P. A., Schreiber S. L., Clardy J. (1993) Atomic structures of the human immunophilin FKBP-12 complexes with FK506 and rapamycin. *J Mol Biol* 229: 105–124.

Vergne I., Chua J., Singh S. B., Deretic V. (2004) Cell biology of *Mycobacterium tuberculosis* phagosome. *Annu Rev Cell Dev Biol* 20: 367–394.

Vergne I., Singh S., Roberts E., Kyei G., Master S., Harris J., de Haro S., Naylor J., Davis A., Delgado M., Deretic V. (2006) Autophagy in immune defense against *Mycobacterium tuberculosis*. *Autophagy* 2: 175–178.

Vetting M. W., Hegde S. S., Javid-Majd F., Blanchard J. S., Roderick S. L. (2002) Aminoglycoside 2'-*N*-acetyltransferase from *Mycobacterium tuberculosis* in complex with coenzyme A and aminoglycoside substrates. *Nat Struct Biol* 9: 653–658.

Vetting M. W., S de Carvalho L. P., Yu M., Hegde S. S., Magnet S., Roderick S. L., Blanchard J. S. (2005) Structure and functions of the GNAT superfamily of acetyltransferases. *Arch Biochem Biophys* 433: 212–226.

Vetting M. W., de Carvalho L. P., Roderick S. L., Blanchard J. S. (2005) A novel dimeric structure of the RimL N^{α}-acetyltransferase from *Salmonella typhimurium*. *J Biol Chem* 280: 22108–22114.

Walburger A., Koul A., Ferrari G., Nguyen L., Prescianotto-Baschong C., Huygen K., Klebl B., Thompson C., Bacher G., Pieters J. (2004) Protein kinase G from pathogenic mycobacteria promotes survival within macrophages. *Science* 304: 1800–1804.

Wei J., Moulder J. W., Roberts E. A., O'Gaora P., Young D. B., Friedman R. L. (2000) Identification of a *Mycobacterium tuberculosis* gene that enhances mycobacterial survival in macrophages. *J Bacteriol* 182: 377–384.

Wei Y., Pattingre S., Sinha S., Bassik M., Levine B. (2008) JNK1-mediated phosphorylation of Bcl-2 regulates starvation-induced autophagy. *Mol Cell* 30: 678–688.

Wirawan E., Vande Walle L., Kersse K., Cornelis S., Claerhout S., Vanoverberghe I., Roelandt R., De Rycke R., Verspurten J., Declercq W., Agostinis P,. Vanden Berghe T., Lippens S., Vandenabeele P. (2010) Caspase-mediated cleavage of Beclin-1 inactivates Beclin-1-induced autophagy and enhances apoptosis by promoting the release of proapoptotic factors from mitochondria. *Cell Death Dis* 1: e18.

Wu S., Barnes P. F., Samten B., Pang X., Rodrigue S,. Ghanny S., Soteropoulos P., Gaudreau L., Howard S. T. (2009) Activation of the eis gene in a W-Beijing strain of *Mycobacterium tuberculosis* correlates with increased SigA levels and enhanced intracellular growth. *Microbiology* 155: 1272–1281.

Zaunbrecher M. A., Sikes R. D. Jr, Metchock B., Shinnick T. M., Posey, J. E. (2009) Overexpression of the chromosomally encoded aminoglycoside acetyltransferase eis confers kanamycin resistance in *Mycobacterium tuberculosis*. *Proc Natl Acad Sci USA* 106: 20004–20009.

국문 초록

결핵균 (*Mycobacterium tuberculosis*)는 결핵을 일으키는 것으로 알려진 대표적 병원성 세균이다. 전세계 인구의 약 1/3 이 결핵균에 감염되어 있으며 결핵으로 인하여 매년 수백만 명이 목숨을 잃는다. 또한 최근에는 기존 항생제에 대한 내성을 가진 multidrug-resistant 와 extensively drug-resistant 한 결핵균들의 등장으로 인하여 그 위험도가 더욱 커지고 있다. 뿐만 아니라 결핵균은 오랜 시간 잠복하고 있다가 인체의 면역력이 떨어졌을 때 활성화되기 때문에 치료를 위하여서는 여러 종류의 항생제를 6-9 개월간 장기 복용하여야 한다. 따라서 더 효과적인 결핵 치료제의 개발이 시급한 상태이다. 특히 결핵균이 면역체계를 이겨 내고 인체 내 세포에서 살아남는다는 것이 곧 결핵균의 병원성과 밀접한 관련이 있다. 그렇기 때문에 결핵균이 대식세포 안에서 어떻게 살아남는 지를 이해하는 것이 매우 중요하다. 결핵균은 숙주세포 내에서 살아남기 위하여 대식세포 내에서 일어나는 phagosomal maturation 과 autophagy 를 방해한다. 이러한 과정에 관여하는 결핵균 단백질의 동정은 새로운 항생제 개발의 표적을 제공해 줄 것이다.

결핵균의 Enhanced intracellular survival (Eis) 단백질을 비병원성인 *Mycobacterium smegmatis* 에서 발현시키면 대식세포 내에서의 생존 능력이 많이 향상된다는 연구 결과가 보고되었다. 결핵균 Eis 가 인체 내 면역체계의 중요한 역할을 하는 proinflammatory cytokine 들의 발현량을 조절하고, JNK pathway 를 통해 autophagy 를 저해 하는 것으로 알려졌다. 또한 결핵균 Eis 가 aminoglycoside 계열의 항생제를 acetylation 시킴으로서 항생제 내성을 유발한다는 보고가 있었다. 이에 본인은 결핵균의 Eis 단백질, 그리고 이와 58%의 아미노산

서열 identity 를 가진 *M. smegmatis* Eis 단백질의 삼차원 구조 및 기능의 비교 연구를 통해 결핵균의 Eis 가 어떻게 알려진 기능들을 하는지 규명하였다.

따라서 결핵균의 Eis 의 다른 기질이 있을 것으로 예상하였고 활성 부위의 구조를 잘 비교하면 결핵균의 Eis 와 *M. smegmatis* 의 Eis 의 기질 부착 부위에 큰 차이가 있음을 발견할 수 있었다. 즉 결핵균의 Eis 의 기질 부착 부위는 좁고 긴 반면, *M. smegmatis* 의 Eis 의 기질 부착 부위는 둥글게 파인 모양을 하고 있다. 따라서 결핵균 Eis 의 기질로서 작용하지만 *M. smegmatis* 의 Eis 의 기질로는 작용하지 못하는 기질이 있을 것으로 판단하고 기존 연구 결과를 토대로 결핵균 Eis 의 기질을 탐색하였다. 그 결과 결핵균 Eis 가 JNK-specific phosphatase 인 dual specificity phosphatase 16 (DUSP16)/MAPK phosphatase-7 (MKP-7)의 kinase 상호작용 부위에 포함되어 있는 Lys55 를 아세틸화 시키는 반면, *M. smegmatis* 의 Eis 는 아세틸화 시키지 못한다는 사실을 알 수 있었다. 추가적인 실험을 통하여 결핵균 Eis 에 의한 아세틸화로 인하여 DUSP16/MKP-7 과 JNK 간의 강화된 상호작용이 JNK 의 autophosphorylation 에 의한 활성화를 저해하고, 결과적으로 인체 대식세포 내 면역체계인 autophagosome 숙성과 autophagy 에 의한 결핵균의 제거가 무력화된다는 모형을 제시하였다. 이러한 결과를 통해 결핵균 Eis 가 새로운 결핵 치료제 개발의 좋은 표적이 될 수 있음을 증명하였다.

Acknowledgements

I would like to express my gratitude to my research supervisor, Professor Se Won Suh, for his advice during the course of these works. I thank Prof. S. B. Kim, Prof. C. M. Park, Prof. J. Y. Lee, and Prof. J. K. Yang for their kind advice to my dissertation. For using synchrotron facilities, I thank the beamline staffs at Pohang Light Source, Korea (BL-4A, BL-6B, and BL-6C) and Photon Factory, Japan (BL-1A, BL-5A, BL-17A, NE3A, and NW12) for assistance during X-ray diffraction experiments.

Appendix：Printouts of first author publications

Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Crystallization and preliminary X-ray crystallographic analysis of a putative agmatinase from *Deinococcus radiodurans*

Ji Ah Lee,‡ Hyung Jun Ahn,‡ Jun Yong Ha, Sun Mi Shim, Kyoung Hoon Kim, Hye Kyung Kim and Se Won Suh*

Department of Chemistry, College of Natural Sciences, Seoul National University, Seoul 151-742, South Korea

‡ These authors contributed equally to this work.

Correspondence e-mail: sewonsuh@snu.ac.kr

Received 22 April 2004
Accepted 2 August 2004

Agmatine, which results from the decarboxylation of arginine by arginine decarboxylase, is a metabolic intermediate in the biosynthesis of putresine and higher polyamines. The enzyme agmatinase catalyses the hydrolysis of agmatine to putresine and urea. Recent studies indicate that agmatinase plays important roles in mammals. Human mitochondrial agmatinase shows a considerable level of sequence similarity to bacterial agmatinases, including a putative agmatinase from *Deinococcus radiodurans*. The putative agmatinase from *D. radiodurans* has been overexpressed in *Escherichia coli* and crystallized at 297 K using polyethylene glycol 3000 as a precipitant. X-ray diffraction data were collected to 1.80 Å from a crystal grown in the presence of Mn^{2+} and 1,6-hexanediamine. The crystals are orthorhombic, belonging to the space group $P2_12_12_1$, with unit-cell parameters a = 81.77, b = 131.44, c = 168.85 Å, $\alpha = \beta = \gamma = 90^\circ$. A hexameric molecule is likely to be present in the asymmetric unit, giving a crystal volume per protein weight (V_M) of 2.15 Å^3 Da^{-1} and a solvent content of 41.8%.

1. Introduction

Agmatine is an amine that is synthesized by decarboxylation of L-arginine by arginine decarboxylase (ADC; Reis & Regunathan, 2000). In bacteria, plants and invertebrates, agmatine is hydrolyzed to putrescine, the precursor of spermine, by agmatinase (agmatine ureohydrolase, EC 3.5.3.11; Reis & Regunathan, 2000). In 1994, agmatine was identified as an endogenous ligand for the imidazoline receptor and agmatinase and ADC were subsequently identified in mammalian brain. Several lines of evidence indicate that agmatinase may play important roles in mammals, including the regulation of neurotransmitter-related functions (Reis & Regunathan, 1998, 2000). The amino-acid sequences of agmatinases indicate their homology to arginases, suggesting a common evolutionary origin (Ouzounis & Kyrpides, 1994; Perozich *et al.*, 1998). Compared with the wealth of structural data on arginases, little structural information is available on agmatinases. The human agmatinase gene encodes a 352-residue protein with a putative mitochondrial targeting sequence at the amino-terminus. Human agmatinase has about 30% sequence identity to bacterial agmatinases and <20% identity to mammalian arginases. Residues required for binding of Mn^{2+} at the active site in bacterial agmatinases and other members of the arginase superfamily are fully conserved in human agmatinase (Mistry *et al.*, 2002; Iyer *et al.*, 2002).

In order to understand the functional differences between agmatinase and arginase in structural terms and to better understand the functional roles of agmatinase, we have initiated three-dimensional structure determination of agmatinase. In this study, we have overexpressed a putative agmatinase from *Deinococcus radiodurans*, a 304-residue protein of 35 751 Da, in *Escherichia coli* and have crystallized it. Of the protein sequences in the SWISS-PROT database, this enzyme shows the highest level of sequence identity to human mitochondrial agmatinase; the sequence identity is 37% over the region encompassing residues 11–298. We report here its crystallization conditions and preliminary X-ray crystallographic data.

2. Experimental

2.1. Protein expression and purification

The putative agmatinase gene (DR0149) was amplified by the polymerase chain reaction using the *D. radiodurans* genomic DNA as template. The forward and reverse oligonucleotide primers designed using the published genome sequence (White *et al.*, 1999) were 5′-G GAA TTC **CAT ATG** AGC GGG CCG GCC CAC C-3′ and 5′-CCG CCG **CTC GAG** GAC ATG GTC GAA CAC CTC GCA C-3′, respectively, where the bases in bold represent the *Nde*I and *Xho*I restriction-enzyme cleavage sites. The amplified DNA was inserted into the *Nde*I/*Xho*I-digested expres-

sion vector pET-28b(+) (Novagen). This vector construction adds an eight-residue tag (LEHHHHHH) to the carboxyl-terminus and a 20-residue tag (MGSSHHHHHHSSGLVPRGSH) to the amino-terminus of the gene product in order to facilitate protein purification. The protein was overexpressed in *E. coli* C41(DE3) cells (Miroux & Walker, 1996). Cells were grown at 310 K to an OD_{600} of 0.5 in Terrific Broth medium containing 50 µg ml^{-1} kanamycin and protein expression was induced by 1.0 m*M* isopropyl-β-D-thiogalactopyranoside (IPTG). Cell growth continued at 293 K for 18 h after IPTG induction and cells were harvested by centrifugation at 4200*g* (6000 rev min^{-1}; Sorvall GSA rotor) for 10 min at 277 K. The cell pellet was resuspended in ice-cold lysis buffer (20 m*M* Tris–HCl pH 7.9, 500 m*M* sodium chloride, 5 m*M* imidazole and 1 m*M* phenylmethylsulfonyl fluoride) and homogenized with an ultrasonic processor. The crude lysate was centrifuged at 70 400*g* (30 000 rev min^{-1}; Beckman 45Ti rotor) for 1 h at 277 K and the recombinant protein in the supernatant fraction was purified in two chromatographic steps. The first step utilized the hexahistidine tags by metal-chelate chromatography on Ni–NTA resin (Qiagen). Next, gel filtration was performed on a HiLoad XK 16 Superdex 200 prep-grade column (Amersham Pharmacia) previously equilibrated with buffer *A* (50 m*M* Tris–HCl pH 7.5) containing 100 m*M* sodium chloride and 1 m*M* manganese sulfate. The homogeneity of the purified protein was assessed by SDS–PAGE (Laemmli, 1970). The protein solution was concentrated using an YM10 ultrafiltration membrane (Amicon). The protein concentration was estimated by measuring the absorbance at 280 nm, employing the calculated extinction coefficient of 16 500 M^{-1} cm^{-1} (SWISS-PROT; http://www.expasy.ch/).

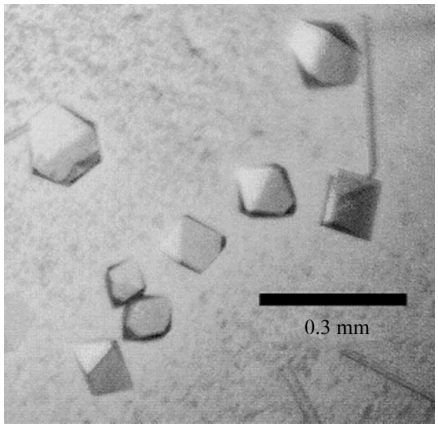


Figure 1
Crystals of the putative agmatinase from *D. radiodurans*. Their approximate dimensions are 0.15 × 0.15 × 0.20 mm.

Table 1
Data-collection statistics.

Values in parentheses refer to the highest resolution shell (1.86–1.80 Å).

X-ray wavelength (Å)	1.000
Temperature (K)	100
Space group	$P2_12_12_1$
Unit-cell parameters (Å)	a = 81.77, b = 131.44, c = 168.85
Resolution range (Å)	30.0–1.80
Total/unique reflections	713493/164504
$R_{\rm merge}$† (%)	7.1 (39.6)
Data completeness (%)	98.0 (94.5)
Average $I/\sigma(I)$	22.4 (3.3)

† $R_{\rm merge} = \sum_h \sum_i |I(h)_i - \langle I(h)\rangle| / \sum_h \langle I(h)\rangle$, where $I(h)i$ is the intensity of the ith measurement of reflection h and $\langle I(h)\rangle$ is the mean value of $I(h)$ for all i measurements.

2.2. Crystallization and dynamic light scattering

Crystallization was performed at 297 K by the hanging-drop vapour-diffusion method using 24-well tissue-culture plates. Each hanging drop was prepared on a siliconized cover slip by mixing equal volumes (6 µl each) of the protein solution at 1.4 m*M* concentration and the inhibitor 1,6-hexanediamine solution at 334 m*M* concentration, resulting in an approximately 1:6.7 molar ratio of the agmatinase monomer to inhibitor. Each hanging drop was placed over 1.0 ml reservoir solution. Initial crystallization conditions were established using screening kits from Hampton Research (Crystal Screens I and II and MembFac) and from deCODE Biostructures Group (Wizard I and II). Dynamic light-scattering experiments were performed using a model DynaPro-801 instrument from Protein Solutions (Lakewood, New Jersey). The data were measured at 297 K with the protein at 1 mg ml^{-1} concentration in 50 m*M* Tris–HCl pH 7.5, 100 m*M* sodium chloride and 1 m*M* manganese sulfate in the presence of a 6.7-fold molar excess of 1,6-hexanediamine.

2.3. X-ray diffraction experiment

Crystals were flash-frozen using a cryoprotectant solution consisting of 100 m*M* sodium phosphate citrate pH 4.2, 200 m*M* sodium chloride, 12%(*w*/*v*) PEG 3000 and 15%(*v*/*v*) glycerol. Crystals were soaked in 5 µl of the cryoprotectant solution for 10 s before being flash-frozen in liquid nitrogen. X-ray diffraction data were collected at 100 K using a DIP-2030 image-plate detector (MacScience) at beamline BL-6B of Pohang Light Source, South Korea. The crystal was rotated through a total of 180° with a 1.0° oscillation range per frame. The wavelength of the synchrotron radiation was 1.000 Å. The raw data were processed and scaled using the *HKL* program package (Otwinowski & Minor, 1997).

3. Results

When the recombinant agmatinase from *D. radiodurans* was expressed in *E. coli* as a fusion with the C-terminal eight-residue tag (LEHHHHHH), the expression level was very low. However, when the 20-residue tag (MGSSHHHHHHSSGLVPRGSH) was also fused at the N-terminus of the protein, it was highly overexpressed in a soluble form, with a much higher yield of ∼40 mg of purified enzyme per litre of culture. Dynamic light-scattering analysis indicated the purified protein to be monodisperse with a polydispersity of 20.5% and a native molecular weight of ∼185 000 Da. This result is consistent with *D. radiodurans* agmatinase existing as a tetramer, pentamer or hexamer in solution (the calculated monomer weight including the N- and C-terminal tags is 35 751 Da).

Despite the presence of both N- and C-terminal tags, the recombinant enzyme readily formed well diffracting crystals. The best crystals were obtained with a reservoir solution comprising 100 m*M* sodium phosphate–citrate pH 4.2, 200 m*M* sodium chloride, 12%(*w*/*v*) PEG 3000 and 30 m*M* glycyl-glycyl-glycine. Rectangular crystals grew to maximum dimensions of 0.15 × 0.15 × 0.20 mm within two weeks (Fig. 1). A flash-frozen crystal diffracted to ∼1.8 Å. A set of diffraction data was collected with a completeness of 98.0% to 1.80 Å and an $R_{\rm merge}$ of 7.1%. Table 1 summarizes the statistics of data collection. The space group was determined to be $P2_12_12_1$ on the basis of systematic absences and the unit-cell parameters are a = 81.77 (4), b = 131.44 (3), c = 168.85 (3) Å, $\alpha = \beta = \gamma = 90°$, where estimated standard deviations are given in parentheses. If it is assumed that a hexameric molecule of the recombinant enzyme is present in the crystallographic asymmetric unit, the crystal volume per protein weight ($V_{\rm M}$) is 2.15 Å^3 Da^{-1} and the solvent content is 41.8% by volume (Matthews, 1968). This assumption is supported by the self-rotation function, which clearly indicated the presence of threefold rotational symmetry. Hence, the hexameric molecule must have 32 symmetry. The structure has been determined by

molecular replacement and the details will be described elsewhere.

We thank Dr H. S. Lee and his staff at beamline BL-6B of Pohang Light Source, South Korea for assistance during synchrotron data collection. This work was supported by a grant from the Korea Ministry of Science and Technology (NRL-2001, grant No. M10318000132). JAL, HJA and KHK are supported by BK21 Fellowships.

References

Iyer, R. K., Kim, H. K., Tsoa, R. W., Grody, W. W. & Cederbaum, S. D. (2002). *Mol. Genet. Metab.* **75**, 209–218.

Laemmli, U. K. (1970). *Nature (London)*, **227**, 680–685.

Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.

Miroux, B. & Walker, J. E. (1996). *J. Mol. Biol.* **260**, 289–298.

Mistry, S. K., Burwell, T. J., Chambers, R. M., Rudolph-Owen, L., Spaltmann, F., Cook, W. J. & Morris, S. M. Jr (2002). *Am. J. Physiol. Gastrointest. Liver Physiol.* **282**, G375–G381.

Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.

Ouzounis, C. A. & Kyrpides, N. C. (1994). *J. Mol. Evol.* **39**, 101–104.

Perozich, J., Hempel, J. & Morris, S. M. Jr (1998). *Biochem. Biophys. Acta*, **1382**, 23–37.

Reis, D. J. & Regunathan, S. (1998). *J. Auton. Nerv. Syst.* **72**, 80–85.

Reis, D. J. & Regunathan, S. (2000). *Trends Pharmacol. Sci.* **21**, 187–193.

White, O. *et al.* (1999). *Science*, **286**, 1571–1577.

Acta Crystallographica Section F
Structural Biology and Crystallization Communications
ISSN 1744-3091

Expression, crystallization and preliminary X-ray crystallographic analyses of two N-terminal acetyltransferase-related proteins from *Thermoplasma acidophilum*

Sang Hee Han,[a]‡ Jun Yong Ha,[a]‡ Kyoung Hoon Kim,[a]‡ Sung Jin Oh,[a] Do Jin Kim,[a] Ji Yong Kang,[a] Hye Jin Yoon,[a] Se-Hee Kim,[b] Ji Hae Seo,[b] Kyu-Won Kim[b] and Se Won Suh[a]*

[a]Department of Chemistry, College of Natural Sciences, Seoul National University, Seoul 151-742, South Korea, and [b]NeuroVascular Coordination Research Center, Research Institute of Pharmaceutical Sciences, College of Pharmacy, Seoul National University, Seoul 151-742, South Korea

‡ These authors contributed equally to this work.

Correspondence e-mail: sewonsuh@snu.ac.kr

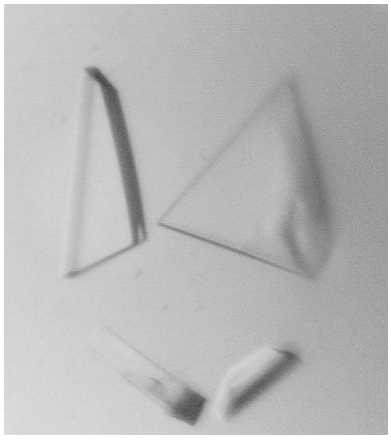
Received 15 September 2006
Accepted 30 September 2006

N-terminal acetylation is one of the most common protein modifications in eukaryotes, occurring in approximately 80–90% of cytosolic mammalian proteins and about 50% of yeast proteins. ARD1 (arrest-defective protein 1), together with NAT1 (*N*-acetyltransferase protein 1) and possibly NAT5, is responsible for the NatA activity in *Saccharomyces cerevisiae*. In mammals, ARD1 is involved in cell proliferation, neuronal development and cancer. Interestingly, it has been reported that mouse ARD1 ($mARD1^{225}$) mediates ε-acetylation of hypoxia-inducible factor 1α (HIF-1α) and thereby enhances HIF-1α ubiquitination and degradation. Here, the preliminary X-ray crystallographic analyses of two N-terminal acetyltransferase-related proteins encoded by the *Ta0058* and *Ta1140* genes of *Thermoplasma acidophilum* are reported. The Ta0058 protein is related to an N-terminal acetyltransferase complex ARD1 subunit, while Ta1140 is a putative N-terminal acetyltransferase-related protein. Ta0058 shows 26% amino-acid sequence identity to both $mARD1^{225}$ and human $ARD1^{235}$.The sequence identity between Ta0058 and Ta1140 is 28%. Ta0058 and Ta1140 were overexpressed in *Escherichia coli* fused with an N-terminal purification tag. Ta0058 was crystallized at 297 K using a reservoir solution consisting of 0.1 *M* sodium acetate pH 4.6, 8%(*w*/*v*) polyethylene glycol 4000 and 35%(*v*/*v*) glycerol. X-ray diffraction data were collected to 2.17 Å. The Ta0058 crystals belong to space group $P4_1$ (or $P4_3$), with unit-cell parameters $a = b = 49.334$, $c = 70.384$ Å, $\alpha = \beta = \gamma = 90^\circ$. The asymmetric unit contains a monomer, giving a calculated crystal volume per protein weight (V_M) of 2.13 Å^3 Da^{-1} and a solvent content of 42.1%. Ta1140 was also crystallized at 297 K using a reservoir solution consisting of 0.1 *M* trisodium citrate pH 5.6, 20%(*v*/*v*) 2-propanol, 20%(*w*/*v*) polyethylene glycol 4000 and 0.2 *M* sodium chloride. X-ray diffraction data were collected to 2.40 Å. The Ta1140 crystals belong to space group *R*3, with hexagonal unit-cell parameters $a = b = 75.174$, $c = 179.607$ Å, $\alpha = \beta = 90$, $\gamma = 120^\circ$. Two monomers are likely to be present in the asymmetric unit, with a V_M of 2.51 Å^3 Da^{-1} and a solvent content of 51.0%.

1. Introduction

N-terminal acetylation is one of the most common protein modifications in eukaryotes, occurring in approximately 80–90% of cytosolic mammalian proteins and about 50% of yeast proteins (Polevoda & Sherman, 2003). Studies with *Saccharomyces cerevisiae* revealed three major N-terminal acetyltransferases, NatA, NatB and NatC, which exhibit different substrate specificities (Polevoda & Sherman, 2003). ARD1 (arrest-defective protein 1), together with NAT1 (*N*-acetyltransferase protein 1), is required for NatA activity in yeast (Lee *et al.*, 1989; Mullen *et al.*, 1989; Park & Szostak, 1992). The yeast N-terminal acetyltransferase complex NatA was quantitatively bound to ribosomes *via* NAT1 and contained a third subunit, NAT5 (Gautschi *et al.*, 2003). Human homologues of ARD1 and NAT1 also form an evolutionarily conserved N-terminal acetyltransferase complex (Arnesen, Anderson *et al.*, 2005) and a human homologue of NAT5 has also been identified (Arnesen, Anderson *et al.*, 2006). In mammals, the N-terminal acetyltransferase complex plays an important role in cell proliferation, neuronal development and cancer (Sugiura *et al.*, 2003; Fisher *et al.*, 2005; Arnesen, Gromyko *et al.*, 2005).

ARD1 and NAT5 proteins have been assigned to the ARD1 subfamily of acetyltransferases. They possess an acetyl-CoA-binding



motif (Q/R)*XX*G*X*(G/A) (Neuwald & Landsman, 1997; Arg82–Ala87 in ARD1 and Arg84–Ala89 in NAT5 for both human and mouse). A recent study has identified three mouse (mARD1^{198}, mARD1^{225} and mARD1^{235}) and two human (hARD1^{131} and hARD1^{235}) variants of ARD1 (Kim *et al.*, 2006). Of the ARD1 isoforms, mARD1^{225} differs completely in its C-terminal region (residues 158–225) from mouse and human ARD1^{235} (Kim *et al.*, 2006). Interestingly, a mouse ARD1 variant (mARD1^{225}) was reported to mediate ε-acetylation of Lys532 within the oxygen-dependent degradation (ODD) domain of hypoxia-inducible factor 1α (HIF-1α), thereby enhancing HIF-1α ubiquitination and degradation (Jeong *et al.*, 2002; Kim *et al.*, 2006). However, other investigators could not demonstrate such an activity for hARD1^{235} (Fisher *et al.*, 2005; Bilton *et al.*, 2005). Overexpression or silencing of hARD1^{235} had no impact on HIF-1α stability (Arnesen, Kong *et al.*, 2005) and the recombinant hARD1^{235} did not catalyze the acetylation of HIF-1α ODD domain (Murray-Rust *et al.*, 2006). In constrast, Chang and coworkers recently provided strong evidence for the role of hARD1 in HIF-1α stability in human cells (Chang *et al.*, 2006). They suggested that connective tissue-growth factor (CTGF) mediated HIF-1α degradation was attributed to hARD1-dependent acetylation of HIF-1α at Lys532 in human lung cancer cells (Chang *et al.*, 2006). Structurally, hARD1^{235} consists of a compact globular region (residues 1–178) and a flexible unstructured C-terminus (Sanchez-Puig & Fersht, 2006). The globular region encompasses the predicted acetyltransferase domain of hARD1^{235} (residues 44–129; Arnesen, Betts *et al.*, 2006). It was also reported that hARD1^{235} forms protofilament aggregates under physiological conditions of pH and temperature (Sanchez-Puig & Fersht, 2006). In particular, the region encompassing the acetyltransferase domain is responsible for fibre formation (Sanchez-Puig & Fersht, 2006), which in turn may abrogate the catalytic activity of hARD1^{235}.

Despite the important functional role of protein N-terminal acetyltransferases, their structural studies are in an early stage. To provide a structural basis to obtain a better understanding of protein N^{α}-acetylation by N-terminal acetyltransferases and to facilitate structural comparisons with N^{ε}-acetyltransferases, we have initiated three-dimensional structure determination of the *Ta0058* gene product from *Thermoplasma acidophilum*. Its 154-residue protein is related to the N-terminal acetyltransferase complex ARD1 subunit. It shows amino-acid sequence identity of 26% with mARD1^{225}, mARD1^{235} and hARD1^{235}. We have also performed crystallization studies of the *Ta1140* gene product from *T. acidophilum*, a putative N-terminal acetyltransferase-related protein. Its 149-residue polypeptide chain shows sequence identity of 28% with Ta0058, but bears no detectable sequence homology with either mouse or human ARD1 variants. As the first step toward structure determination of the Ta0058 and Ta1140 proteins from *T. acidophilum*, we overexpressed them in *Escherichia coli* and crystallized them. Here, we report the crystallization conditions and preliminary X-ray crystallographic data. Both of these crystals are suitable for structure determination at sufficiently high resolution. It is hoped that the ultimate structural information on these proteins will provide a framework for obtaining a better insight into the molecular functions of N-terminal acetyltransferase-related proteins.

2. Experimental

2.1. Protein expression and purification

The *Ta0058* and *Ta1140* genes were amplified by polymerase chain reaction (PCR) using the genomic DNA of *T. acidophilum* as the template and the following oligonucleotide primers, which were designed using the published genome sequence (Ruepp *et al.*, 2000). The forward and reverse primers were 5′-G GAA TTC **CAT ATG** GCT ATA AAT GCT GTG GCC G-3′ and 5′-CCG CCG **CTC GAG** TTA CTA AAC TAT GCG CCA CAT AGT GT A-3′ for Ta0058 and 5′-G GAA TTC **CAT ATG** ATC CTG AGA AGA TAC AGA AGT ACG G-3′ and 5′-CCG **CTC GAG** TTA TCT GCA CAT CAA CCT CAT TCT ATA-3′ for Ta1140. The bases in bold represent the *Nde*I and *Xho*I restriction-enzyme cleavage sites. The PCR product was then digested with *Nde*I and *Xho*I and inserted into the *Nde*I/*Xho*I-digested expression vector pET-28b(+) (Novagen). This vector construction added a 20-residue tag (MGSSHHHHHHSSGLVPR-GSH) to the N-terminus of the gene product in order to facilitate protein purification.

The Ta0058 protein was expressed in *E. coli* Rosetta II(DE3)pLysS cells. The cells were grown at 310 K to an OD_{600} of ∼0.5 in Terrific Broth medium containing 30 µg ml^{-1} kanamycin and protein expression was induced by the auto-induction method (Studier, 2005). Cell growth continued at 303 K for 18 h after auto-induction and cells were harvested by centrifugation at 4200*g* (6000 rev min^{-1}; Sorvall GSA rotor) for 10 min at 277 K. The cell pellet was resuspended in ice-cold lysis buffer [5 m*M* imidazole, 500 m*M* sodium chloride, 20 m*M* Tris–HCl pH 7.9, 5%(*v*/*v*) glycerol, 1 m*M* phenylmethylsulfonyl fluoride, 0.3%(*v*/*v*) Tween-20] and was then homogenized with an ultrasonic processor. The crude cell extract was centrifuged at 36 000*g* (18 000 rev min^{-1}; Hanil Supra 21K rotor) for 60 min at 277 K and the recombinant protein in the supernatant fraction was purified in two chromatographic steps. The first step utilized the N-terminal hexahistidine tag by metal-chelate chromatography on Ni–NTA resin (Qiagen). Gel filtration was then performed on a HiLoad XK 16 Superdex 200 prep-grade column (Amersham-Pharmacia) previously equilibrated with buffer *A* (20 m*M* Tris–HCl pH 7.9) containing 200 m*M* sodium chloride. The homogeneity of the purified protein was assessed by SDS–PAGE. The protein solution was concentrated using a YM10 ultrafiltration membrane (Millipore-Amicon). The protein concentration was estimated by measuring the absorbance at 280 nm.

The Ta1140 protein was expressed in *E. coli* BL21(DE3)pLysS cells. The cells were grown at 310 K to an OD_{600} of ∼0.6 in Terrific Broth medium containing 30 µg ml^{-1} kanamycin and protein expression was induced by 0.5 m*M* isopropyl β-D-thiogalactopyranoside (IPTG). Cell growth continued at 293 K for 26 h after IPTG induction and cells were harvested by centrifugation at 4200*g* (6000 rev min^{-1}; Sorvall GSA rotor) for 10 min at 277 K. The cell pellet was resuspended in ice-cold lysis buffer [5 m*M* imidazole, 500 m*M* sodium chloride, 20 m*M* Tris–HCl pH 7.9, 5%(*v*/*v*) glycerol, 1 m*M* phenylmethylsulfonyl fluoride] and was then homogenized with an ultrasonic processor. The crude cell extract was centrifuged at 36 000*g* (18 000 rev min^{-1}; Hanil Supra 21K rotor) for 60 min at 277 K and the recombinant protein in the supernatant fraction was purified in two chromatographic steps, essentially as above, except that the gel-filtration buffer contained 2 m*M* β-mercaptoethanol and its pH was 7.0.

2.2. Crystallization

Crystallization trials were carried out at 297 K using the hanging-drop vapour-diffusion method with 24-well VDX plates (Hampton Research). Each hanging drop was prepared by mixing equal volumes (2 µl each) of the Ta0058 protein solution (at 7.1 mg ml^{-1} concentration in a buffer consisting of 20 m*M* Tris–HCl pH 7.9, 200 m*M* sodium chloride) and the reservoir solution. Each hanging drop was

Table 1
Data-collection statistics for Ta0058.

Values in parentheses are for the highest resolution shell.

X-ray source	Photon Factory beamline BL-5A
X-ray wavelength (Å)	1.0000
Temperature (K)	100
Space group	$P4_1$ (or $P4_3$)
Unit-cell parameters (Å, °)	a = 49.334, b = 49.334, c = 70.384, α = 90, β = 90, γ = 90
Resolution range (Å)	50–2.17 (2.25–2.17)
Total/unique reflections	61931/8885
$R_{\rm merge}$† (%)	0.070 (0.182)
Data completeness (%)	99.0 (92.2)
Average $I/\sigma(I)$	28.5 (5.56)

† $R_{\rm merge} = \sum_h \sum_i |I(h)_i - \langle I(h)\rangle| / \sum_h \sum_i I(h)_i$, where $I(h)_i$ is the intensity of the ith measurement of reflection h and $\langle I(h)\rangle$ is the mean value of $I(h)$ for all i measurements.

Table 2
Data-collection statistics for Ta1140.

Values in parentheses are for the highest resolution shell.

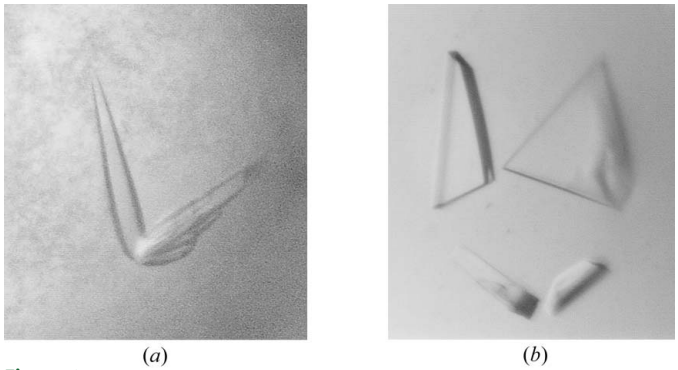
X-ray source	Pohang Light Source beamline BL-4A
X-ray wavelength (Å)	0.9919
Temperature (K)	100
Space group	$R3$
Unit-cell parameters (hexagonal setting) (Å, °)	a = 75.174, b = 75.174, c = 179.607, α = 90, β = 90, γ = 120
Resolution range (Å)	50–2.40 (2.49–2.40)
Total/unique reflections	74875/15318
$R_{\rm merge}$† (%)	0.053 (0.478)
Data completeness (%)	99.9 (99.7)
Average $I/\sigma(I)$	31.8 (3.26)

† $R_{\rm merge}$ as in Table 1.

prepared by mixing equal volumes (2 µl each) of the Ta1140 protein solution (at 6 mg ml^{-1} concentration in a buffer consisting of 20 m*M* Tris–HCl pH 7.0, 2 m*M* β-mercaptoethanol and 200 m*M* sodium chloride) and the reservoir solution. It was placed over 0.45 ml reservoir solution. Screening of crystallization conditions was carried out using commercial kits from Hampton Research (Crystal Screens I and II, MembFac, Index I and II and SaltRX I and II) and from Emerald Biostructures Inc. (Wizard I and II).

2.3. X-ray diffraction experiments

The reservoir solution used to crystallize Ta0058, which consisted of 0.1 *M* sodium acetate pH 4.6, 8%(w/*v*) polyethylene glycol 4000 and 35%(*v*/*v*) glycerol, served as a cryoprotectant solution. Therefore, the crystals could be flash-frozen by dipping into liquid nitrogen. Ta1140 crystals were cryoprotected using a solution consisting of 95 m*M* trisodium citrate pH 5.6, 19%(*v*/*v*) 2-propanol, 19%(*w*/*v*) polyethylene glycol 4000 and 5%(*v*/*v*) glycerol. The crystal was briefly (∼10 s) soaked in the cryoprotectant solution and was flash-frozen by dipping into liquid nitrogen. For Ta0058, we collected X-ray diffraction data at 100 K using a Quantum 315 CCD detector (Area Detector System Corporation, Poway, CA, USA) at beamline BL-5A of Photon Factory, Japan. The crystal was rotated through a total of 180° with a 1.0° oscillation range per frame. For Ta1140, X-ray data were collected at 100 K using a Quantum 210 CCD detector (Area Detector System Corporation, Poway, CA, USA) at beamline BL-4A of Pohang Light Source, Korea. The crystal was rotated through a total of 152°. The raw data for both Ta0058 and Ta1140 were processed and scaled using the *HKL*-2000 program package (Otwinowski & Minor, 1997).



(*a*) (*b*)

Figure 1
Crystals of N-terminal acetyltransferase-related proteins from *T. acidophilum*. (*a*) Crystals of Ta0058. The large crystals have approximate dimensions of 0.4 × 0.05 × 0.05 mm. (*b*) Crystals of Ta1140. The large crystals have approximate dimensions of 0.2 × 0.2 × 0.02 mm.

3. Results

When we expressed Ta0058 in *E. coli* strain BL21(DE3) as an intact form, it was not expressed. When we expressed it as a fusion with the N-terminal 20-residue tag (MGSSHHHHHHSSGLVPRGSH), it was highly overexpressed but the protein was insoluble. When the N-terminally tagged Ta0058 was expressed in Rosetta II(DE3)pLysS cells, it was highly overexpressed and was mostly soluble. The yield was ∼11 mg purified enzyme per litre of culture. When Ta1140 was expressed as an intact form in the *E. coli* BL21(DE3)pLysS strain, the protein was expressed at a very low level with poor solubility (below 10%). However, when we expressed it as a fusion with the N-terminal 20-residue tag in BL21(DE3)pLysS cells, the expression level was much higher and ∼80% of the expressed protein was in the soluble fraction. The yield was ∼8 mg of purified enzyme per litre of culture. Despite the presence of the N-terminal tag, we could obtain well diffracting crystals of both recombinant proteins.

The best crystallization conditions of Ta0058 were obtained after optimizing the initial crystallization condition No. 37 of Crystal Screen I (Hampton Research), with the refined reservoir solution comprising 0.1 *M* sodium acetate pH 4.6, 8%(*w*/*v*) polyethylene glycol 4000 and 35%(*v*/*v*) glycerol. The crystals grew to approximate dimensions of 0.4 × 0.05 × 0.05 mm within a day (Fig. 1). The crystal used for data collection diffracted strongly to 2.17 Å, but only very weakly beyond this resolution. An X-ray diffraction data set was collected to 2.17 Å and the crystals belonged to space group $P4_1$ (or $P4_3$), with unit-cell parameters $a = b = 49.334$, $c = 70.384$ Å, $\alpha = \beta = \gamma = 90°$. One monomer is present in the crystallographic asymmetric unit, with a calculated crystal volume per protein weight ($V_{\rm M}$) of 2.13 Å^3 Da^{-1} and a solvent content of 42.1%. Table 1 summarizes the X-ray data-collection statistics for Ta0058.

We grew the best crystals of Ta1140 after optimizing the initial crystallization condition No. 40 of Crystal Screen I (Hampton Research), with the refined reservoir solution comprising 0.1 *M* trisodium citrate pH 5.6, 20%(*v*/*v*) 2-propanol, 20%(*w*/*v*) polyethylene glycol 4000 and 0.2 *M* sodium chloride. The crystals grew to approximate dimensions of 0.2 × 0.2 × 0.02 mm within 2 d (Fig. 1). The crystals of Ta1140 diffracted more poorly than those of Ta0058. Diffraction data were collected to 2.40 Å and the space group was determined to be $R3$ on the basis of systematic absences and diffraction intensity symmetry. The unit-cell parameters are $a = b = 75.174$, $c = 179.607$ Å, $\alpha = \beta = 90$, $\gamma = 120°$ in the hexagonal setting. Two monomers are likely to be present in the asymmetric

unit, giving a V_M of 2.51 Å^3 Da^{-1} and a solvent content of 51.0%. Table 2 summarizes the X-ray data-collection statistics for Ta1140.

We thank the staff of beamlines BL-5A, BL-6A and NW-12A of Photon Factory, Tsukuba, Japan and beamlines BL-4A and BL-6B of Pohang Light Source, Korea for assistance during X-ray experiments. This work was supported by a grant from the Center for Functional Analysis of Human Genome to SWS and by the Creative Research Initiatives (NeuroVascular Coordination Research Center) of MOST/KOSEF, Korea to KWK. SHH, JYH, KHK, OSJ, DJK and JYK are recipients of BK21 fellowships.

References

Arnesen, T., Anderson, D., Baldersheim, C., Lanotte, M., Varhaug, J. E. & Lillehaug, J. R. (2005). *Biochem. J.* **386**, 433–443.
Arnesen, T., Anderson, D., Torsvik, J., Halseth, H. B., Varhaug, J. E. & Lillehaug, J. R. (2006). *Gene*, **371**, 291–295.
Arnesen, T., Betts, M. J., Pendino, F., Liberles, D. A., Anderson, D., Caro, J., Kong, X., Varhaug, J. E. & Lillehaug, J. R. (2006). *BMC Biochem.* **7**, 13.
Arnesen, T., Gromyko, D., Horvli, O., Fluge, O., Lillehaug, J. & Varhaug, J. E. (2005). *Thyroid*, **15**, 1131–1136.
Arnesen, T., Kong, X., Evjenth, R., Gromyko, D., Varhaug, J. E., Lin, Z., Sang, N., Caro, J. & Lillehaug, J. R. (2005). *FEBS Lett.* **579**, 6428–6432.
Bilton, R., Mazure, N., Trottier, E., Hattab, M., Déry, M. A., Richard, D. E., Pouysségur, J. & Brahimi-Horn, M. C. (2005). *J. Biol. Chem.* **280**, 31132–31140.
Chang, C. C., Lin, M. T., Lin, B. R., Jeng, Y. M., Chen, S. T., Chu, C. Y., Chen, R. J., Chang, K. J., Yang, P. C. & Kuo, M. L. (2006). *J. Natl Cancer Inst.* **98**, 984–995.
Fisher, T. S., Etages, S. D., Hayes, L., Crimin, K. & Li, B. (2005). *J. Biol. Chem.* **280**, 17749–17757.
Gautschi, M., Just, S., Mun, A., Ross, S., Rucknagel, P., Dubaquie, Y., Ehrenhofer-Murray, A. & Rospert, S. (2003). *Mol. Cell Biol.* **23**, 7403–7414.
Jeong, J. W., Bae, M. K., Ahn, M. Y., Kim, S. H., Sohn, T. K., Bae, M. H., Yoo, M. A., Song, E. J., Lee, K. J. & Kim, K. W. (2002). *Cell*, **111**, 709–720.
Kim, S. H., Park, J. A., Kim, J. H., Lee, J. W., Seo, J. H., Jung, B. K., Chun, K. H., Jeong, J. W., Bae, M. K. & Kim, K. W. (2006). *Biochem. Biophys. Res. Commun.* **340**, 422–427.
Lee, F. J., Lin, L. W. & Smith, J. A. (1989). *J. Bacteriol.* **171**, 5795–5802.
Mullen, J. R., Kayne, P. S., Moerschell, R. P., Tsunasawa, S., Gribskov, M., Colavito-Shepanski, M., Grunstein, M., Sherman, F. & Sternglanz, R. (1989). *EMBO J.* **8**, 2067–2075.
Murray-Rust, T. A., Oldham, N. J., Hewitson, K. S. & Schofield, C. J. (2006). *FEBS Lett.* **580**, 1911–1918.
Neuwald, A. F. & Landsman, D. (1997). *Trends Biochem. Sci.* **22**, 154–155.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Park, E. C. & Szostak, J. W. (1992). *EMBO J.* **11**, 2087–2093.
Polevoda, B. & Sherman, F. (2003). *Biochem. Biophys. Res. Commun.* **308**, 1–11.
Ruepp, A., Graml, W., Santos-Martinez, M. L., Koretke, K. K., Volker, C., Mewes, H. W., Frishman, D., Stocker, S., Lupas, A. N. & Baumeister, W. (2000). *Nature (London)*, **407**, 508–513.
Sanchez-Puig, N. & Fersht, A. R. (2006). *Protein Sci.* **15**, 1968–1976.
Studier, F. W. (2005). *Protein Expr. Purif.* **41**, 207–234.
Sugiura, N., Adams, S. M. & Corriveau, R. A. (2003). *J. Biol. Chem.* **278**, 40113–40120.

Chapter 11

From No Expression to High-Level Soluble Expression in *Escherichia coli* by Screening a Library of the Target Proteins with Randomized N-Termini

Kyoung Hoon Kim, Jin Kuk Yang, Geoffrey S. Waldo, Thomas C. Terwilliger, and Se Won Suh

For structural studies by x-ray crystallography and nuclear magnetic resonance it is important for the target protein to be available in large quantity and high purity. *Escherichia coli* expression systems remain the most versatile and convenient means to produce a large quantity of recombinant proteins. Unfortunately, some proteins fail to be expressed in *E. coli* or are expressed in an insoluble form. To overcome the difficulty of no expression or expression at a very low level, a simple and efficient approach of screening a library of variants of a target protein with randomized N-termini was devised. In this method, a few N-terminal residues are randomized by designing a mixture of oligonucleotides for the forward PCR primer and fuse the library in front of green fluorescent protein, which serves as a reporter for the target protein expression level and folding yield. In favorable cases this approach can result in high-level soluble expression of recombinant proteins in *E. coli*. This chapter describes the results of a test of this approach with a bacterial protein (the HI0952 gene product) that is not well expressed in *E. coli*.

1. Introduction

Production of properly folded proteins or their isolated domains is a prerequisite for the detailed structural and functional characterization. For instance, structural studies of proteins by x-ray crystallography or nuclear magnetic resonance normally require significant amounts (5–50 mg) of a target protein in high purity. This is most often accomplished by expression of the target protein in soluble form using a suitable heterologous host. *Escherichia coli* is still the first choice for protein expression as it offers many advantages over other expression hosts in terms of speed, cost effectiveness, and convenience. Many expression vectors for heterologous expression in *E. coli* are already available and new vectors continue to be developed.

From: *Methods in Molecular Biology, Vol. 426: Structural Proteomics: High-throughput Methods*
Edited by: B. Kobe, M. Guss and T. Huber © Humana Press Inc., Totowa, NJ

E. coli has some limitations as an expression host. One of the limitations is that some proteins fail to be expressed in *E. coli* or their level of expression or solubility is very low when they are expressed. Causes of such failure in protein expression in *E. coli* may include the toxicity of the target protein, improper protein folding, the codon usage problem, instability of the mRNA, susceptibility of the expressed protein to proteolysis, and the hairpin formation at the 5'-side of the mRNA. Several strategies are available to overcome some of these difficulties. Rosetta cells (Novagen) provide tRNAs corresponding to infrequently used codons. C41 and C43 cells, derivatives of BL21, were isolated for producing toxic proteins *(1)*. mRNA secondary structures at the 5'-side can be minimized by silent mutations of the gene *(2)*.

Alterations at the N-terminus of the target protein are well known to affect heterologous protein expression *(3–6)*. It was observed that NGG codons at positions +2, +3, and +5 downstream of the initiation codon lower the gene expression in *E. coli* at the translational level and it has been suggested that the low expression is not the result of mRNA secondary structure or a lowered intracellular mRNA pool *(7)*. As a specific example of altering the N-terminal region for enhancing the expression level, sequences within the first seven amino acid codons of the eukaryotic membrane protein (bovine cytochrome P450 17α-hydroxylase) were altered to optimize the expression in *E. coli (8)*. However, such methods are time consuming and usually require a large number of trials.

Here a simple and efficient approach was devised to reducing the problem of no expression or very low-level expression in *E. coli*. In this method, few N-terminal residues of the target protein were randomized by employing a mixture of oligonucleotides as the 5'-primer for polymerase chain reaction (PCR) and fuse the library in front of green fluorescent protein, which is used as the reporter for the expression level and folding yield. Then the library of the target protein with randomized N-terminus is screened for the clones that express the target protein at high levels in a soluble form. In favorable cases this approach has the potential to enable expression of the recombinant protein in *E. coli* both at a sufficiently high level and in a soluble form. This approach was tested with a bacterial protein (HI0952 gene product) that was not well expressed in *E. coli*. GFP fusions can be expressed in a variety of cell types, so in principle this method is applicable to protein expression in other hosts, provided that methods for highly efficient transformation of host cells are available.

2. Materials

2.1. Enzymes and Proteins

1. T4 DNA ligase (Takara, Shiga, Japan)
2. NdeI (DCC Bionet, Sungnam, Korea)
3. BamHI (DCC Bionet, Sungnam, Korea)
4. High-fidelity thermophilic DNA polymerase (Phusion DNA polymerase from Finnzymes, Keilaranta, Finland)

2.2. Plasmids, Nucleotides, Gels, and Kits

1. Green fluorescent protein (GFP) folding reporter vector (X-FR) *(9)*
2. pET-21a(+) (Merck, Darmstadt, Germany)

3. Oligonucleotide primers (Bioneer, Daejon, Korea) (see Note 1)
4. Template DNA for PCR: For cloning the HI0952 gene, the genomic DNA was obtained from ATCC
5. 1% (w/v) Agarose gel (Bioworld, Dublin, OH)
6. 12.5% (w/v) SDS-PAGE gel
7. Qiaquick spin gel extraction kit (Qiagen, Hilden, Germany)

2.3. Cells and Media

1. *E. coli* strain ElectroMAX DH10B (Invitrogen, Grand Island, NY)
2. *E. coli* strain BL21(DE3) (Invitrogen)
3. *E. coli* strain Rosetta II(DE3)pLysS (Invitrogen)
4. Luria-Bertani (LB) medium (BD Sciences, Franklin Lakes, NJ)
5. Kanamycin stock solution (35 mg/ml): 1.75 g of kanamycin in 50 ml of sterile water. Aliquot in 1 ml fractions and store at −20°C
6. Isopropyl-β-D-thio-galactopyranoside (IPTG) stock solution (1.0 M): 4.8 g IPTG in 20 ml of sterile water. Aliquot in 1 ml fractions and store at −20°C
7. SOC: 2% Bacto tryptone, 0.5% Bacto yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$, 20 mM glucose

2.4. Equipment

1. Microcentrifuge
2. PCR machine (MJ Research Model PTC-200)
3. MicroPulser electroporation apparatus (Bio-Rad, Hercules, CA)
4. Gene Pulser cuvette: 0.2 cm electrode gap (Bio-Rad, catalog no. 165–2086)
5. Nitrocellulose membrane, polyester reinforced (GE Osmonics, Minnetonka, MN): 85 mm, 137 mm diameter
6. Kirby-Bauer plate: 90 mm, 150 mm diameter
7. Illumatool tunable lights system, 488 nm excitation filter and 520 nm emission filter (Lightools Research, Encinitas, CA)
8. Digital camera with a stand

3. Methods

3.1. Library Construction of the Target Protein with Randomized N-Terminus

1. For the forward primer, design a mixture of oligonucleotides that contain NNY for the positions +2, +3, and +4 downstream of the translation initiation codon, where N is any base (A, T, G, or C) and Y is a pyrimidine base (C or T) (see Notes 1 and 2).
2. Amplify the target gene by running the PCR reaction (see Note 3).
3. The PCR product is digested with NdeI and BamHI.
4. The digested PCR product is inserted into the NdeI/BamHI-digested GFP folding reporter vector by T4 DNA ligase.

3.2. Amplification of Library Plasmid

The GFP folding reporter vector containing the library of the target protein with randomized N-terminus is transformed into the *E. coli* nonexpression

strain DH10B by the electroporation method (see Note 4) and the amount of the plasmid library is amplified. The amplified plasmid library is then isolated for transformation into the *E. coli* expression strain BL21(DE3) by the chemical method. The following protocols for library transformation and recovery of the amplified library are a slight modification of those already described by Waldo *(10)*.

3.2.1. Library Transformation into DH10B by Electroporation

1. Thaw 100 µl aliquot of DH10B cells on ice (10–15 min).
2. Prechill 100 µl dd H_2O on ice (10 min) and three Gene Pulser cuvettes per library.
3. Combine 100 µl thawed DH10B and 50 µL pre-chilled dd H_2O, add the diluted DH10B to same tube on ice containing the 10 µL resuspended, ligated DNA, flick to gently mix, incubate 5 min on ice.
4. Transfer 50 µL of transformation mix in each Gene Pulser cuvette (see Note 4).
5. Recover each of the three transformations by immediately resuspending in 1 ml SOC in 12 ml culture tube with shaking at 37°C for 1.5 hours.
6. Centrifuge the recovered cultures in 1.5-ml Eppendorf tubes for 1 min at 14,000 rpm.
7. Leave 200 µL of supernatant over the pelleted cells (remove ~ 800 µl supernatant). Resuspend cells by pipetting.
8. Pool the three tubes into one 1.5-ml Eppendorf. The total volume will be about 600 µl.
9. Plate the library onto a large Kirby-Bauer plate (150 mm) selective media LB plate supplemented with 35 µg/ml kanamycin (LB-Kan).
10. On a separate counting agar plate, plate 1/300th of library (2 µl of the 600 µl pooled transformations into pool of ~ 200 µl SOC onto a standard LB/Kan selective plate to get the counts). Incubate 12–16 hours at 37°C (see Note 5).

3.2.2. Library Plasmid Recovery

1. Add 12 ml LB to the Kirby-Bauer DH10B library plate from plates in Section 3.2.1., resuspend with spreader.
2. Transfer suspension to 15-ml Falcon tube, vortex to suspend.
3. Perform QIAgen plasmid prep on 750 µl of cell suspension. Cell mass prepped should be equivalent to 3 ml of overnight LB culture, i.e., approx. 75–100 mg pellet.

3.3. Screening of the Clones on the Basis of Gfp Fluorescence

1. Prepare nitrocellulose membranes (137 mm).
2. Label then wet the membranes (see Note 6). Place on towels to dry until damp only.
3. Lay on an LB-Kan Kirby-Bauer plate (150 mm), excluding all air pockets.
4. Incubate at 37°C for at least 1 hour prior to plating.
5. The plasmid library (1 µl) is transformed into BL21(DE3) competent cells (50 µl) by the heat shock protocol (42°C, 1 min) (see Note 7).
6. The transformed cells are plated directly onto nitrocellulose membranes (137 mm) on an LB-agar plate (150 mm). After the membrane is dry, invert the plate, and incubate at 37°C for 10–15 hours until colonies are approximately 1 mm in diameter (see Note 7).

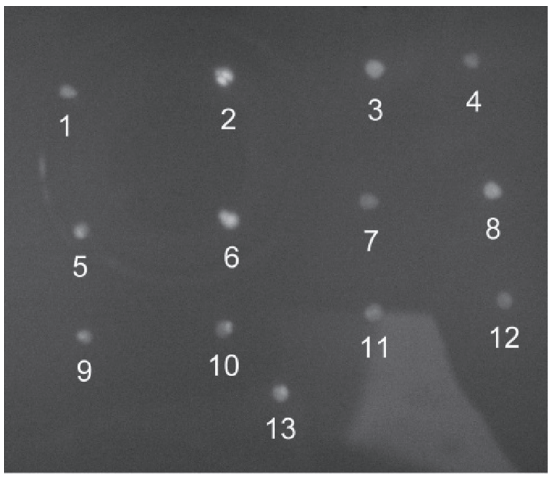


Fig. 11.1 Colonies of the selected HI0952 variants showing significant fluorescence improvements over the wild type, whose GFP fusion is almost nonfluorescent. Among these 13 colonies, colonies #2, #3, #5, #6, #8, and #13 were chosen for further characterization. The wild type colony is not included in this figure, because it is invisibly dark.

7. After incubation, the membrane is transferred onto an LB-agar plate (150 mm) containing 2 mM IPTG and is incubated for 5–6 hours for induction.
8. Pick the 30 brightest colonies and transfer them onto the master LB-agar plate (90 mm).
9. The master LB-agar plate (90 mm) is incubated at 37°C for 14–16 hours and its replica is made on a nitrocellulose membrane (85 mm).
10. The replica membrane is incubated on an LB-agar plate (90 mm) at 37°C for 8–10 hours, transferred onto an LB-agar plate containing 2 mM IPTG, and incubated for an additional 4–6 hours for induction.
11. Select colonies showing significant fluorescence improvements over the wild-type and the cell mass of the selected colonies on the plate is recovered. For HI0952, we selected 13 colonies (Fig. 11.1).

3.4. Test of Protein Expression as a Fusion with GFP

1. The selected variant colonies are overexpressed in BL21(DE3) cells.
2. Cells are grown at 37°C up to OD_{600} of 0.5 in LB medium containing 35 μg/ml kanamycin.
3. The protein expression is induced by 1.0 mM IPTG.
4. Cells are grown for additional 18 hours at 30°C.
5. The expression levels of the GFP fusion protein for the selected colonies are estimated by running SDS-PAGE on the total *E. coli* proteins.
6. Choose a few colonies that have a reasonably high expression level. As an example, six colonies were chosen for HI0952 (colonies #2, #3, #5, #6, #8, and #13 in Fig. 11.1).
7. The solubility of the GFP fusion protein for the selected colonies is estimated by running SDS-PAGE on the soluble and insoluble fractions of the total *E. coli* proteins (see Note 8).

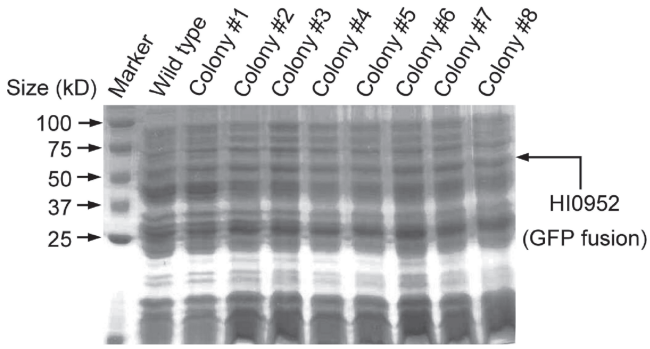


Fig. 11.2 SDS-PAGE analysis of HI0952 variants with C-terminally fused GFP. The colony #3 shows both bright fluorescence and a high expression level.

8. Choose a colony that has both a high degree of soluble expression and a reasonably high expression level. As an example, the colony #3 was chosen for HI0952 (Fig. 11.2).
9. Purify the GFP folding reporter plasmid containing the desired variant.

3.5. Test of Protein Expression Without GFP Fusion

1. The variant gene is amplified by high-fidelity PCR using the purified GFP plasmid as the template and the appropriate primers (see Note 9) for subcloning into an expression vector without GFP fusion.
2. The amplified PCR product is digested with NdeI and BamHI.
3. The digested PCR product is inserted into the pET-21a(+) expression vector, which has been already digested with the same restriction enzymes.
4. The plasmid pET-21a(+) carrying the desired variant gene is transformed into BL21(DE3) cells by the standard calcium chloride method.
5. The transformed cells are grown at 37°C up to OD_{600} of 0.5 in LB medium containing 35 μg ml^{-1} kanamycin.
6. The protein expression is induced by 1.0 mM IPTG.
7. Cells are grown for additional 18 hours at 30°C.
8. The expression level and solubility of the variant protein are assessed by running SDS-PAGE on soluble and insoluble fractions of the total *E. coli* protein. An example of SDS-PAGE analysis is shown in Fig. 11.3.
9. The mutations introduced to the N-terminus are determined by performing DNA sequencing (see Note 10).
10. If necessary, the conditions for soluble expression of the variant protein can be further optimized, for example, by expressing it in different *E. coli* cells (e.g., BL21(DE3)pLysS or Rosetta II (DE3)pLysS) or at lower temperatures (e.g., 15°C). The variant protein can be evolved for further improvement in solubility using the split-GFP *in vivo* solubility reporter system *(11)*.

4. Notes

1. Primer sequences designed for PCR (sub)cloning of the HI0952 gene are as follows. The underlined bases represent NdeI and BamHI digestion sites. The rare codon CGG was changed for Arg2 into CGT in the forward primer #1b for the wild-type to reduce the codon usage problem.

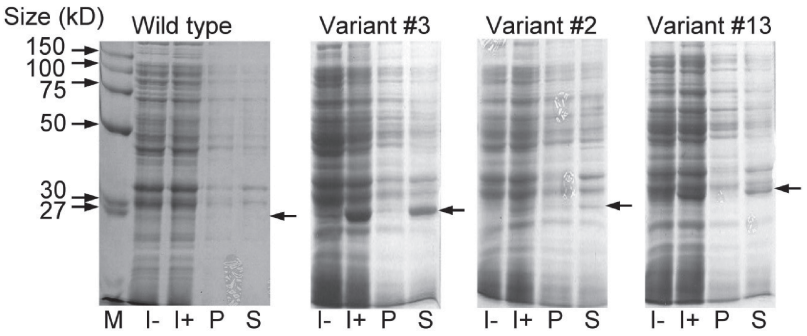


Fig. 11.3 SDS-PAGE analysis of HI0952 wild-type and variants (nonfusion), which are indicated by the arrows on the right side of each gel. The #3 variant shows the best solubility compared to other variants and the wild-type. I–, before IPTG induction; I+, after IPTG induction; P, pellet fraction; S, supernatant fraction.

- #1a (Forward primer for constructing a library of proteins with randomized N-terminus):
  G GAA TTC CAT ATG NNY NNY NNY TCT TTC TTA TTT TTC TTT TAT AAA TAT ATG (52-mer), where N is any base (A, T, G, or C) and Y is a pyrimidine base (C or T).
- #1b (Forward primer for wild type):
  G GAA TTC CAT ATG CGT TCA AAA TCT TTC TTA TTT TTC TTT TAT AAA TAT ATG (52-mer).
- #2 (Reverse primer for both wild type and the variant library; reverse primer for subcloning the selected variant into pET-21a(+)):
  CGC CGC GGA TCC TAA TAA ACA ATT CTC CGC GAA AGA A (37-mer).
- #3 (Forward primer for subcloning the selected variant into pET-21a(+)):
  CTA ATA CGA CTC ACT ATA GGG G (22-mer). This primer covers a T7 promoter region.

2. By limiting the third position of the codon to be either C or T we can avoid introducing stop codons and some of the rare codons in *E. coli*. That is, we exclude all codons for five amino acids (methionine, glutamine, glutamate, lysine, and tryptophan) and stop codons by limiting the codon to NNY. The designed primer mixture contains $(4 \times 4 \times 2)^3 = 32{,}768$ sequences, which correspond to $(15 \times 15 \times 15) = 3{,}375$ protein sequences. By limiting the codon to NNY, we also exclude the NGG codons at positions +2 and +3, which were found to lower the gene expression in *E. coli* at the translational level *(7)*. It is reasonable to assume that the structure and function of the recombinant protein will be affected only insignificantly by altering the second, third, and fourth amino acids in most cases. This is because the altered positions are generally variable in sequence.
3. For HI0952, we used the following protocol ("touchdown" PCR) programmed into a MJ Research Model PTC-200 PCR machine:

   Step 1: 98°C, 30 seconds
   Step 2: 98°C, 10 seconds

Step 3: 50°C, 20 seconds (Increase the temperature by 1°C per cycle.)
Step 4: 72°C, 20 seconds*
Step 5: Steps 2 through 4 are repeated 15 times
Step 6: 98°C, 10 seconds
Step 7: 58°C, 20 seconds
Step 8: 72°C, 20 seconds*
Step 9: Steps 6–8 are repeated 30 times.
Step 10: 72°C, 5 minutes
Step 11: 4°C

Details of the PCR reaction mixture:

35 µl ddH2O
10 µl 5 × High-fidelity buffer (Finnzymes)
1.0 µl 10 mM dNTPs (2.5 mM each dNTP)
0.5 µl Phusion DNA polymerase (Finnzymes)
0.5 µl Forward primer (100 pmol/µl stock)
0.5 µl Reverse primer (100 pmol/µl stock)
1.0 µl Template DNA (50 ng/µl)
1.5 µl DMSO
50 µl Total volume

4. Set electroporator to 2.5 kV. If arcing occurs and is outside the cuvette, recover mix and perform pulse in a fresh cuvette.
5. If fewer than 200 colonies are observed on the counting plate, the variant library may not be as complete as desired. In such a case, try to adjust the ratio of the GFP folding reporter plasmid to the PCR product to obtain more than 200 colonies.
6. It is necessary to preincubate the nitrocellulose membrane in distilled water for 5 minutes in order to avoid formation of air bubbles between the membrane and the plate, which adversely affect the growth of the colonies.
7. Depending on the competency of the BL21(DE3) strain, 1 µl of a "standard" plasmid prep could yield several hundred thousand colonies. One option is to make the BL21(DE3) expression library essentially a lawn, then washing the viable cells off the plate and diluting sequentially twice 350-fold and plating 1 ml of this per Kirby-Bauer plate (150 mm) for actually picking. This would add one extra step per day. Otherwise, one could do a set of dilution plates in step 6 of Heading 3.3, and use the one showing single colonies for picking. Since the library size is relatively small, this places an estimate on the number of clones needed to get good representation (see Note 5).
8. Fluorescence of the C-terminal GFP in the fusion context reflects the expression level of the fusion protein as well as the lack of misfolding by the upstream fusion partner protein. Therefore, increased fusion fluorescence is correlated with increased expression and solubility of the test protein in the nonfusion context. Note that GFP-fusions can have reduced solubility, so it is essential to test the solubility of the selected protein variants without the fused GFP.

*For Phusion DNA polymerase (Finnzymes), the supplier recommends an extension time of 15 seconds per 1 kb for low complexity DNAs (e.g., plasmid, lambda, or BAC DNA) and 30 seconds per 1 kb for high-complexity genomic DNAs. A longer extension time than recommended by the supplier is preferred for this PCR reaction.

9. Forward primer: #3 primer (*see* Note 1). Reverse primer: #2 primer (see Note 1).
10. Results of DNA sequencing of some variants of HI0952 are as follows.

    Wild-type: ATG CGT TCA AAA (MRSK)
    Colony #2: ATG ACC TGT GAC (MTCD)
    Colony #3: ATG CGC TGT TGC (MRCC)
    Colony #8: ATG GGC CTT GCC (MGLA)
    Colony #13: ATG GGT CTT GAT (MGLD)

Acknowledgments

This work was supported by the Korea Ministry of Science and Technology (NRL-2001).

References

1. Miroux, B., and Walker, J. E. (1996) Over-production of proteins in *Escherichia coli*: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. J. *Mol. Biol*. **260**, 289–298.
2. Cèbe, R., and Geiser, M. (2006) Rapid and easy thermodynamic optimization of the 5'-end of mRNA dramatically increases the level of wild type protein expression in *Escherichia coli*. *Protein Expr. Purif*. **45**, 374–380.
3. Kis, M., Burbridge, E., Brock, I. W., Heggie, L., Dix, P. J., and Kavanagh, T. A. (2004) An N-terminal peptide extension results in efficient expression, but not secretion, of a synthetic horseradish peroxidase gene in transgenic tobacco. *Ann. Bot. (Lond)* **93**, 303–310.
4. Orchard, S. S., and Goodrich-Blair, H. (2005) An encoded N-terminal extension results in low levels of heterologous protein production in *Escherichia coli*. *Microb. Cell. Fact*. **21**, 22–31.
5. Sati, S. P., Singh, S. K., Kumar, N., and Sharma, A. (2002) Extra terminal residues have a profound effect on the folding and solubility of a *Plasmodium falciparum* sexual stage-specific protein over-expressed in *Escherichia coli*. *Eur. J. Biochem*. **269**, 5259–5263.
6. Sawant, S. V., Kiran, K., Singh, P. K., and Tuli, R. (2001) Sequence architecture downstream of the initiator codon enhances gene expression and protein stability in plants. *Plant Physiol*. **126**, 1630–1636.
7. Gonzalez de Valdivia, E. I., and Isaksson, L. A. (2004) A codon window in mRNA downstream of the initiation codon where NGG codons give strongly reduced gene expression in *Escherichia coli*. *Nucleic Acids Res*. **32**, 5198–5205.
8. Barnes, H. J., Arlotto, M. P., and Waterman, M. R. (1991) Expression and enzymatic activity of recombinant cytochrome P450 17α-hydroxylase in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **88**, 5597–5601.
9. Waldo, G. S., Standish, B. M., Berendzen, J., and Terwilliger, T. C. (1999) Rapid protein-folding assay using green fluorescent protein. *Nature Biotechnol*. **17**, 691–695.
10. Waldo, G. S. (2003) Improving protein folding efficiency by directed evolution using the GFP folding reporter. *Methods Mol. Biol*. **230**, 343–359.
11. Cabantous, S., Pedelacq, J. D., Mark, B. L., Naranjo, C., Terwilliger, T. C., and Waldo, G. S. (2005) Recent advances in GFP folding reporter and split-GFP solubility reporter technologies. Application to improving the folding and solubility of recalcitrant proteins from *Mycobacterium tuberculosis*. *J. Struct. Funct. Genomics* **6**, 113–119.

PROTEINS
STRUCTURE ■ FUNCTION ■ BIOINFORMATICS

STRUCTURE NOTE

Crystal structure of FAF1 UBX domain in complex with p97/VCP N domain reveals a conformational change in the conserved FcisP touch-turn motif of UBX domain

Kyoung Hoon Kim,[1] Wonchull Kang,[2] Se Won Suh,[1,3]* and Jin Kuk Yang[2]*

[1] Department of Chemistry, College of Natural Sciences, Seoul National University, Seoul 151-747, Korea

[2] Department of Chemistry, College of Natural Sciences, Soongsil University, Seoul 156-743, Korea

[3] Department of Biophysics and Chemical Biology, College of Natural Sciences, Seoul National University, Seoul 151-747, Korea

Key words: Fas-associated factor 1; AAA+ family ATPase; adaptor protein binding.

INTRODUCTION

p97/valosin-containing protein (VCP) is an ATPase of the AAA+ family and has been implicated in diverse cellular processes: post-mitotic Golgi reassembly, endoplasmic reticulum associated degradation, nuclear envelope reconstruction, cell cycle, suppression of apoptosis, and DNA damage response.[1] AAA+ family members are classified into two subgroups of Type I and Type II on the basis of the number of ATPase domains included. Type II contains two ATPase domains, referred to D1 and D2 domains, and Type I contains only one ATPase domain, termed D2. In addition to the ATPase domains, less conserved domains are often found at the N-terminus (N domain) or the C-terminus of many AAA+ family proteins, and those terminal domains are involved in binding to adaptor proteins. p97/VCP belongs to the Type II subgroup and is comprised of the N, D1, and D2 domains, respectively. Structural studies have shown that p97/VCP forms a homo-hexamer through D1 and D2 domains, and the N domain is located at the perimeter of the hexameric ring.[2–4]

p97/VCP accomplishes its various functions through recruiting different adaptor proteins. Proteins containing the UBX domain such as Fas-associated factor 1 (FAF1) and p47 form the largest group of the adaptors.[5,6] The UBX domain is a general p97/VCP-binding module,[7] and its structures were determined by NMR spectroscopy on FAF1 UBX and p47 UBX.[6,8] The UBX domain adopts a β-grasp fold similar to ubiquitin and its conserved FP sequence motif in the loop connecting strands S3 and S4 (so called S3/S4 loop) is inserted into the cleft of the p97/VCP N domain as shown in the crystal structure of p47 UBX in complex with p97/VCP ND1 fragment at 2.9 Å resolution.[9] Very recently, we reported the crystal structure of FAF1 UBX, revealing that the very S3/S4 loop containing the FP sequence adopts a cis-Pro touch-turn structure, a rare case of open β-turn, and the peptide bond between these two residues is a cis peptide, so named as FcisP touch-turn motif.[10]

FAF1, initially identified as a Fas-interacting protein, inhibits the proteasome-mediated protein degradation process either by interacting with p97/VCP which may serve as a molecular chaperone presenting the ubiquitinated client proteins to the proteasome or by interacting with ubiquitinated client proteins.[5] Human FAF1, com-

Additional Supporting Information may be found in the online version of this article.

Abbreviations: FAF1, Fas-associated factor 1; VCP, valosin-containing protein.

Kyoung Hoon Kim and Wonchull Kang contributed equally to this work.
Grant sponsor: Korea Healthcare Technology R&D Project, Ministry of Health and Welfare, Republic of Korea; Grant number: A080930
*Correspondence to: Jin Kuk Yang, Department of Chemistry, College of Natural Sciences, Soongsil University, Seoul 156-743, Korea. E-mail: jinkukyang@ssu.ac.kr (or) Se Won Suh, Department of Chemistry, College of Natural Sciences, Seoul National University, Seoul 151-747, Korea. E-mail: sewonsuh@snu.ac.kr
Received 11 March 2011; Revised 13 April 2011; Accepted 27 April 2011
Published online in Wiley Online Library (wileyonlinelibrary. com).
DOI: 10.1002/prot.23073

posed of 650 amino acid residues, contains five domains: UBA at the N-terminus, two tandem UBLs, UAS, and UBX at the C-terminus.[5,11] Among those five, the two very terminal domains, UBA and UBX, are implicated in proteasome-mediated protein degradation. The N-terminal UBA domain interacts with ubiquitins conjugated to the client proteins and the C-terminal UBX domain binds to the N domain of p97/VCP. Through these two kinds of interactions implicated in the proteasome-mediated protein degradation, FAF1 inhibits the protein degradation and consequently enhance the cell death.[5]

In an effort to elucidate the conserved structural features of the UBX domain as a general p97/VCP-binding module and also to provide a structural basis of the interaction between FAF1 and p97/VCP, we have carried out X-ray crystallographic structure analyses of FAF1 UBX domain in complex with p97/VCP N domain. Here, we report its crystal structure at 2.2 Å resolution, revealing conformational changes of FAF1 UBX domain on binding.

MATERIALS AND METHODS

Data collection and structure determination

The human FAF1 UBX domain (residues 571–650) and the human p97/VCP N domain (residues 21–196) were overexpressed, purified, and crystallized as described previously.[12,13] We collected X-ray diffraction data at 2.2 Å resolution at Photon Factory, Japan on the beamline BL-17A. The structure was solved by the molecular replacement using Phaser in CCP4 program suite.[14,15] The search model for FAF1 UBX domain was its previously reported NMR structure (PDB code 1H8C).[8] For the p97/VCP N domain, the corresponding part (residues 23–196) was taken from the crystal structure of p97/VCP ND1 in complex with p47 UBX (PDB code 1S3S).[9] Domains were located in the order of p97/VCP N domain and FAF1 UBX. Subsequent manual model rebuilding was carried out using Coot[16] interspersed with rounds of automatic refinement by Refmac5[17] and CNS.[18] Geometry of the final refined model was checked with MolProbity.[19] Statistics for data collection and structure refinement is summarized in Table I. Structure factors and the final refined coordinates have been deposited in the Protein Data Bank with an accession number 3QC8.

Site-directed mutagenesis

Four single mutants (R579A, F619S, R621A, and F645S) and two double mutants (F619S/R621A and F619S/F645S) of FAF1 UBX were prepared using the method of the QuikChange Site-Directed Mutagenesis kit (Stratagene, USA). The mutations were confirmed by sequencing (Bionics, Korea).

Table I
Data Collection and Refinement Statistics

Structure	FAF1 UBX in complex with p97/VCP N
Data collection	
Space group	$C222_1$
Unit cell parameters	58.07 Å; 72.74 Å; 130.59 Å
Resolution	30.0–2.20 Å
No. of measured reflections	204,811
No. of unique reflections	14,425
R_{merge} [a]	7.9% (28.9%)
Completeness	99.9% (100.0%)
$<I/\sigma(I)>$	4.7 (2.6)
Redundancy	14.2 (14.3)
Refinement	
R_{work}/R_{free} [b]	21.7%/25.0%
r.m.s.d. bond lengths	0.024 Å
r.m.s.d. bond angles	1.72°
Subunits in asymmetric unit	1 + 1
No. of atoms (protein/water)	2019/130
Average B factors (protein/water)	46.7 Å^2/50.7 Å^2
Ramachandran allowed region	99.8%
Ramachandran disallowed region	0.4%

Values in parentheses are for the highest resolution shell of 2.32–2.20 Å.
[a] $R_{merge} = \Sigma_h\Sigma_i|I(h)_i - <I(h)>|/\Sigma_h\Sigma_i I(h)_i$, where $I(h)$ is the intensity of reflection h, Σ_h is the sum over all reflections, and Σ_i is the sum over i measurements of reflection h.
[b] $R = \Sigma_h||F_{obs}(h)| - |F_{calc}(h)||/\Sigma_h|F_{obs}(h)|$, where $F(h)$ is the structure factor of reflection h. R_{free} was calculated using 5% data excluded from refinement, and R_{work} using the other 95% data.

Isothermal titration calorimetry

Isothermal titration calorimetry (ITC) experiments were performed using a VP-ITC instrument (MicroCal, USA). Recombinant protein samples were dialyzed into 50 m*M* Tris-HCl (pH 7.4), 150 m*M* KCl, 5 m*M* $MgCl_2$, and 10% (v/v) glycerol before titration. Titrations were carried out by injecting consecutive aliquots of FAF1 UBX wild type or mutants (1 m*M*) protein sample into the ITC cell containing p97/VCP N domain sample (0.05 m*M*) at 25°C. Binding stoichiometry, enthalpy, entropy, and binding constants were determined by fitting the data to a one-site binding model. The ITC data were fit using Origin 7.0 (MicroCal, USA).

RESULTS AND DISCUSSION

Structure of FAF1 UBX in complex with p97/VCP N domain

The asymmetric unit of the crystal contains a single copy of 1:1 complex between FAF1 UBX domain and p97/VCP N domain. The refined model of the complex contains residues Glu571 to Glu650 of FAF1 UBX and residues Pro23 to Glu192 of p97/VCP N domain. FAF1 UBX domain comprises five β-strands (S1–S5), one α-helix (H1), and three 3_{10} helices (G1–G3) [Fig. 1(A)]. The five strands form a mixed β-sheet in the order of 21534, and the sheet rolls around the α-helix H1 to form

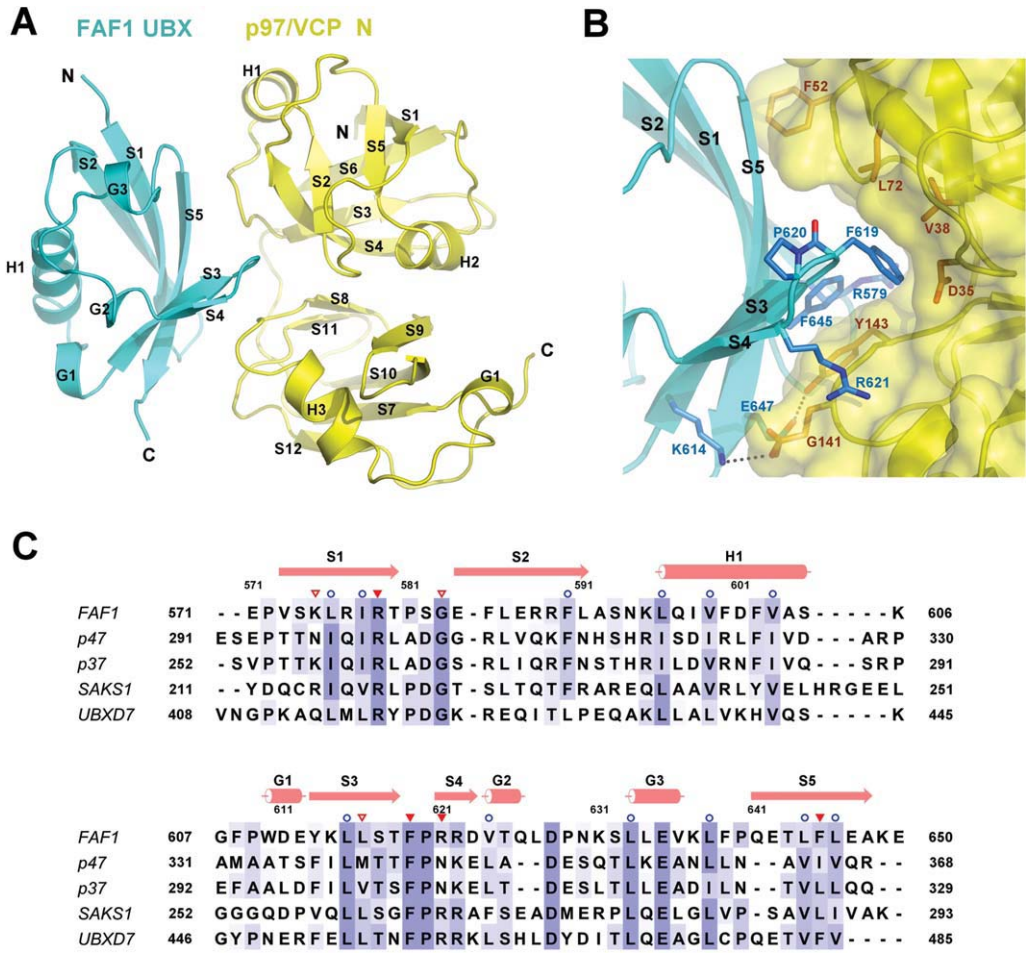


Figure 1

Structure FAF1 UBX in complex with p97/VCP N domain. **A**: Overall structure of the complex between FAF1 UBX and p97/VCP N domain. **B**: Closeup view of the binding interface. Several key residues are shown and the dashed line indicates a hydrogen bond. Transparent surface is presented for p97/VCP N domain. **C**: Sequence alignment of UBX domains from various proteins. These UBX domains have been experimentally shown to bind to p97/VCP N domain. Alignment was carried out by ClustalX[20] and edited by Jalview.[21] Blue shading represents the conservation. The red triangle is for the residues in the binding interface and the blue circle for the residues forming the hydrophobic core of the fold.

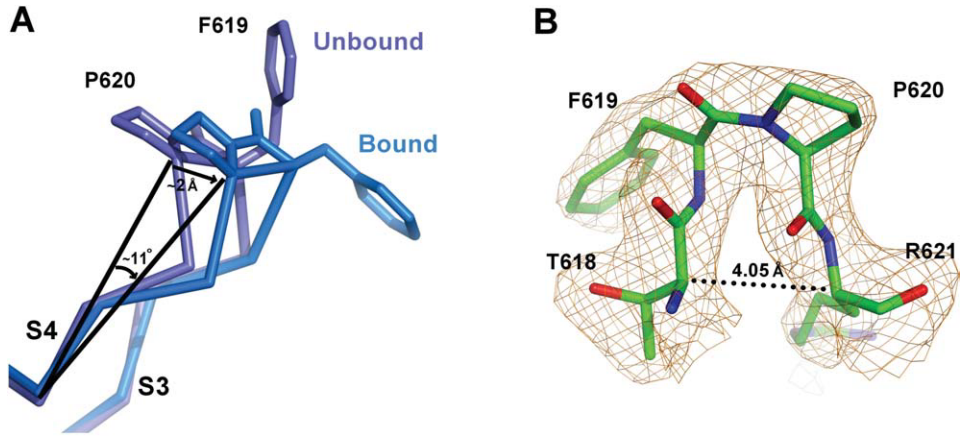


Figure 2

FcisP touch-turn. **A**: Conformational change of the FcisP touch-turn, or S3/S4 loop. The "unbound" is from the our recent crystal structure of free FAF1 UBX and the "bound" is the current structure of FAF1 UBX in complex with p97/VCP N domain. **B**: FcisP touch-turn, a rare case of an open β-turn, formed by the four residues TFPR (residues 618–621). The dashed line represents the distance between $C\alpha^{i}$ and $C\alpha^{i+3}$.

the well-known β-grasp fold, as previously shown in its NMR and crystal structures.[8,10]

It should be noted that the cis peptide between F^{619} and P^{620} of FAF1 UBX was observed for the first time in our recently reported 2.9 Å crystal structure of FAF1 UBX alone[10] and is confirmed by the high-resolution electron density of the current 2.2 Å structure (Fig. 2). In the previous NMR structures of UBX domains from p47 and FAF1 and the crystal structure of p47 UBX domain in complex with p97/VCP ND1, the same peptide bond was described commonly as trans.[6,8,9] This point should be emphasized because the conserved FcisP touch-turn motif containing these two residues is a signature of p97/VCP-binding UBX domains and plays a critical role in their binding. It is also noteworthy that the FcisP touch-turn motif in the S3/S4 loop clearly shows the characteristics of a cis-Pro touch-turn, a rare case of an open β-turn. The distance between $C\alpha^{i}$ and $C\alpha^{i+3}$ and dihedral angles φ and ψ of the F^{619} match very well with the defined values of the cis-Pro touch-turn; the distance is 4.05 Å, which is the range of the defined 4–5 Å, and the dihedral angles are (+46, +65) for F^{619} which fall into the defined (+60, +60) region [Fig. 2(B)].[22] Therefore, our finding of the conserved FcisP touch-turn motif in UBX domain from its recently reported crystal structure is further validated by the current high-resolution structure in complex with p97/VCP N domain.

The overall structure of p97/VCP N domain has two characteristic subdomains (Nn subdomain of residues 25–106 and Nc of residues 112–186) forming a dumbbell-shape as previously observed [Fig. 1(A)].[2] The main β-sheet of FAF1 UBX domain faces one side of the dumbbell-shaped p97/VCP N domain and the entire buried surface area covers about 1680 $Å^2$ which is in the typical range for protein–protein interactions.[23] Between the two subdomains of p97/VCP N domain is formed a cleft into which FcisP touch-turn of FAF1 UBX is inserted [Fig. 1(A)]. The side chain phenyl ring of F^{619} on the turn is deeply inserted into the cleft and makes hydrophobic contacts with three residues (D^{35}, V^{38}, and L^{72}) of p97/VCP N domain. Another phenylalanine F^{645} on the strand S5 makes hydrophobic contacts through stacking with Y^{143} at the bottom of the cleft. Surrounding those two phenylalanines (F^{619} and F^{645}), several hydrophilic residues, including R^{579}, K^{614}, R^{621}, and E^{647}, form hydrogen bonds to residues of p97/VCP near the cleft. The two arginine residues, R579 and R621, that flank phenylalanine residues F619 and F645 together with cis-P620 line up and insert into the cleft [Fig. 1(B)]. These five residues including cis-P^{620} are highly conserved among the UBX-containing proteins with experimentally shown binding activity toward p97/VCP N domain [Fig. 1(C)]. This structural observation, together with the sequence conservation, implies that the sequence motif composed of R^{579}-F^{619}cis$P^{620}R^{621}$/N-F^{645}/L/I/V, must be a signature of the p97/VCP-binding activity of UBX domains.

To assess the contributions of these conserved residues in the motif to the binding activity, we prepared four single-point mutants (R579A, F619S, R621A, and F645S) and two double-point mutants (F619S-R621A and F619S-F645S). Isothermal calorimetric titration was carried out with the purified proteins p97/VCP N domain and these FAF1 UBX mutants or the wild type (Supporting Information Fig. S1). The binding stoichiometry was determined to be 1 (measured 1.06) indicating that FAF1 UBX and p97/VCP N domain binds in a 1:1 ratio. The dissociation constant, K_d, for the wild type FAF1 UBX was measured to be 11.2 μ*M*. The single mutations of R579A, F619S, R621A, or F645S increased K_d by an order of magnitude, whereas the double mutations of F619S-R621A or F619S-F645S by two orders of magnitude (Supporting Information Table S1). Among these four residues, mutation of F^{619} exerted the strongest effect. It is also notable that any single-point mutation is not enough to abolish the interaction. The double-point mutants (F619S-R621A and F619S-F645S) show 95–237 times greater K_d than the wild type and they seem to have lost a greater part of the binding activity. Conclusively, the conserved interface residues of R^{579}-F^{619}cis$P^{620}R^{621}$/N-F^{645}/L/I/V motif of FAF1 UBX domain are important for its binding activity, as confirmed by isothermal calorimetric titrations and as expected from the crystal structure.

Conformational change in FcisP touch-turn motif of UBX domain

The p97/VCP N domain bound to FAF1 UBX in our crystal structure does not show any noticeable structural difference from its free state observed in the crystal structure of ND1, Ref. 2, (PDB ID 1E32) with the an r.m.s. deviation of 0.7 Å for the superposed Cα atoms of residues 24–191. It implies that the p97/VCP N domain is rigid enough not to undergo any significant conformational change upon binding to FAF1 UBX domain. In contrast, the FAF1 UBX bound to p97/VCP N domain shows an important structural difference in its FcisP touch-turn motif (or S3/S4 loop region, residues 618–621) compared with its unbound state that was recently observed in our recently reported crystal structure of FAF1 UBX alone.[10] Upon binding to p97/VCP N domain, the Cα atoms of residues F^{619} and P^{620} move by about 11° in angle and about 2 Å in distance toward the cleft in the p97/VCP N domain [Fig. 2(A)]. Interestingly, the side chain phenyl ring of F^{619} partly stacks over P^{620} in the unbound state, whereas it swings out in the bound state toward the cleft in p97/VCP N domain and forms hydrophobic interactions with D^{35}, V^{38}, and L^{72} in the cleft [Fig. 2(A)]. These structural differences of FAF1 UBX between its unbound and bound states imply that its binding to p97/VCP N domain induces a conformational change in the region of FcisP touch-turn motif,

resulting in insertion of the turn into the cleft in p97/VCP N domain. This kind of conformational change was not previously described for the UBX domain of p47, another p97/VCP-binding UBX protein, whose structure was determined in its unbound state using NMR spectroscopy[6] and in a complex with p97/VCP ND1 using X-ray crystallography.[9]

It is noteworthy that the binding stoichiometry of the FAF1-p97/VCP complex is an open question. Our crystal structure and the isothermal calorimetric titration data indicate that FAF1 UBX domain binds to p97/VCP N domain in a 1:1 ratio. However, it does not necessarily mean that the full-length FAF1 and p97/VCP forms a 6:6 complex. This notion is corroborated by the fact that p47 UBX domain can bind to p97/VCP in a 6:6 ratio but the full-length p47 binds to p97/VCP to form a 3:6 complex.[24,25]

In summary, we determined the crystal structure of FAF1 UBX domain in complex with p97/VCP N domain. It revealed that a significant conformational change of FAF1 UBX domain is induced around the conserved FcisP touch-turn upon its insertion into the cleft of p97/VCP N domain for binding. The FcisP touch-turn moves by about 11° (~2 Å in distance) toward the cleft, with the side-chain phenyl ring of the phenylalanine residue swinging out and getting deep into the cleft. In addition, mutational studies, together with the structural observations, indicate that five highly conserved residues, R^{579}-F^{619}cis$P^{620}R^{621}$/N-F^{645}/L/I/V, are important for the binding of FAF1 UBX domain to p97/VCP N domain.

ACKNOWLEDGMENTS

The authors thank the staff of the Pohang Light Source beamlines 4A in Korea, and the Photon Factory beamlines BL17A in Japan.

REFERENCES

1. Vij N. AAA ATPase p97/VCP: cellular functions, disease and therapeutic potential. J Cell Mol Med 2008;12:2511–2518.
2. Zhang X, Shaw A, Bates PA, Newman RH, Gowen B, Orlova E, Gorman MA, Kondo H, Dokurno P, Lally J, Leonard G, Meyer H, van Heel M, Freemont PS. Structure of the AAA ATPase p97. Mol Cell 2000;6:1473–1484.
3. DeLaBarre B, Brunger AT. Complete structure of p97/valosin-containing protein reveals communication between nucleotide domains. Nat Struct Biol 2003;10:856–863.
4. DeLaBarre B, Brunger AT. Nucleotide dependent motion and mechanism of action of p97/VCP. J Mol Biol 2005;347:437–452.
5. Song EJ, Yim SH, Kim E, Kim NS, Lee KJ. Human Fas-associated factor 1, interacting with ubiquitinated proteins and valosin-containing protein, is involved in the ubiquitin-proteasome pathway. Mol Cell Biol 2005;25:2511–2524.
6. Yuan X, Shaw A, Zhang X, Kondo H, Lally J, Freemont PS, Matthews S. Solution structure and interaction surface of the C-terminal domain from p47: a major p97-cofactor involved in SNARE disassembly. J Mol Biol 2001;311:255–263.
7. Schuberth C, Buchberger A. UBX domain proteins: major regulators of the AAA ATPase Cdc48/p97. Cell Mol Life Sci 2008;65: 2360–2371.
8. Buchberger A, Howard MJ, Proctor M, Bycroft M. The UBX domain: a widespread ubiquitin-like module. J Mol Biol 2001;307:17–24.
9. Dreveny I, Kondo H, Uchiyama K, Shaw A, Zhang X, Freemont PS. Structural basis of the interaction between the AAA ATPase p97/VCP and its adaptor protein p47. EMBO J 2004;23:1030–1039.
10. Kang W, Yang, J.K. Crystal structure of human FAF1 UBX domain reveals a novel FcisP touch-turn motif in p97/VCP-binding region. Biochem Biophys Res Commun 2011;407:531–534.
11. Menges CW, Altomare DA, Testa JR. FAS-associated factor 1 (FAF1): diverse functions and implications for oncogenesis. Cell Cycle 2009;8:2528–2534.
12. Shin HY, Kang W, Lee SY, Yang JK. Crystallization and preliminary X-ray crystallographic analysis of N domain of p97/VCP in complex with UBX domain of FAF1. Acta Crystallogr 2010;F66:41–43.
13. Kang W, Shin HY, Yang JK. Crystallization and preliminary X-ray crystallographic analysis of human FAF1 UBX domain. Acta Crystallogr 2010;F66 (Part 2):211–213.
14. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. Phaser crystallographic software. J Appl Crystallogr 2007;40 (Part 4):658–674.
15. Collaborative Computational Project Number 4. The CCP4 suite: programs for protein crystallography. Acta Crystallogr Sect D 1994;D50 (Part 5):760–763.
16. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. Acta Crystallogr Sect D 2004;D60 (Part 12):2126–2132.
17. Vagin AA, Steiner RA, Lebedev AA, Potterton L, McNicholas S, Long F, Murshudov GN. REFMAC5 dictionary: organization of prior chemical knowledge and guidelines for its use. Acta Crystallogr Sect D 2004;D60 (Part 12):2184–2195.
18. Brunger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS, Read RJ, Rice LM, Simonson T, Warren GL. Crystallography and NMR system: a new software suite for macromolecular structure determination. Acta Crystallogr Sect D 1998;D54 (Part 5):905–921.
19. Davis IW, Leaver-Fay A, Chen VB, Block JN, Kapral GJ, Wang X, Murray LW, Arendall WB, III, Snoeyink J, Richardson JS, Richardson DC. MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. Nucleic Acids Res 2007;35 (Web Server issue):W375–W383.
20. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG. Clustal W and Clustal X version 2.0. Bioinformatics 2007;23:2947–2948.
21. Waterhouse AM, Procter JB, Martin DM, Clamp M, Barton GJ. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. Bioinformatics 2009;25:1189–1191.
22. Videau LL, Arendall WB, III, Richardson JS. The cis-Pro touch-turn: a rare motif preferred at functional sites. Proteins 2004;56: 298–309.
23. Lo Conte L, Chothia C, Janin J. The atomic structure of protein-protein recognition sites. J Mol Biol 1999;285:2177–2198.
24. Beuron F, Dreveny I, Yuan X, Pye VE, McKeown C, Briggs LC, Cliff MJ, Kaneko Y, Wallis R, Isaacson RL, Ladbury JE, Matthews SJ, Kondo H, Zhang X, Freemont PS. Conformational changes in the AAA ATPase p97-p47 adaptor complex. EMBO J 2006;25:1967–1976.
25. Kondo H, Rabouille C, Newman R, Levine TP, Pappin D, Freemont P, Warren G. p47 is a cofactor for p97-mediated membrane fusion. Nature 1997;388:75–78.

THE JOURNAL OF BIOLOGICAL CHEMISTRY VOL. 287, NO. 14, pp. 10727–10737, March 30, 2012

Helical Repeat Structure of Apoptosis Inhibitor 5 Reveals Protein-Protein Interaction Modules*[S]

Received for publication, October 27, 2011, and in revised form, February 9, 2012 Published, JBC Papers in Press, February 12, 2012, DOI 10.1074/jbc.M111.317594

Byeong-Gu Han[‡§1], Kyoung Hoon Kim[¶1], Sang Jae Lee[‡1], Kyung-Chae Jeong[‡], Jea-Won Cho[‡], Kyung Hee Noh[‖], Tae Woo Kim[‖], Soon-Jong Kim[**], Hye-Jin Yoon[¶], Se Won Suh[¶], Sangho Lee[§], and Byung Il Lee[‡2]

*From the [‡]Biomolecular Function Research Branch, Division of Convergence Technology, Research Institute, National Cancer Center, Goyang, Gyeonggi 410-769, Korea, the [¶]Department of Chemistry and Department of Biophysics and Chemical Biology, College of Natural Sciences, Seoul National University, Seoul 151-747, Korea, the [‖]Department of Biomedical Sciences, Graduate School of Medicine, Korea University, Seoul 136-701, Korea, the **Department of Chemistry, Mokpo National University, Chonnam 534-729, Korea, and the [§]Department of Biological Sciences, Sungkyunkwan University, Suwon, Gyeonggi 440-746, Korea*

Background: Up-regulated in various cancers, API5 prevents apoptosis under growth factor deprivation.
Results: We have determined the crystal structure of API5 with the HEAT and ARM repeat and show that Lys-251 acetylation is important for its function.
Conclusion: API5 likely serves as a scaffold for multiprotein complex with its cellular function regulated by lysine acetylation.
Significance: Structural basis of API5 function is important in targeting anti-apoptosis.

Apoptosis inhibitor 5 (API5) is an anti-apoptotic protein that is up-regulated in various cancer cells. Here, we present the crystal structure of human API5. API5 exhibits an elongated all α-helical structure. The N-terminal half of API5 is similar to the HEAT repeat and the C-terminal half is similar to the ARM (Armadillo-like) repeat. HEAT and ARM repeats have been implicated in protein-protein interactions, suggesting that the cellular roles of API5 may be to mediate protein-protein interactions. Various components of multiprotein complexes have been identified as API5-interacting protein partners, suggesting that API5 may act as a scaffold for multiprotein complexes. API5 exists as a monomer, and the functionally important heptad leucine repeat does not exhibit the predicted a dimeric leucine zipper. Additionally, Lys-251, which can be acetylated in cells, plays important roles in the inhibition of apoptosis under serum deprivation conditions. The acetylation of this lysine also affects the stability of API5 in cells.

Apoptosis, a programmed cell death process, plays important roles in sculpting the developing organism and in maintaining cell number homeostasis (1, 2). It is also critical for effective cancer chemotherapy (3). Two main pathways of apoptosis have been studied extensively. The extrinsic pathway is mediated by interactions between the death receptors and death ligands. This induces the formation of the death-inducing signaling complex and caspase-8 (4). The intrinsic pathway, which is triggered by radiation, drugs, reactive oxygen species, and radicals, begins with the release of cytochrome *c* from mitochondria, which ultimately activates caspase-9 (4, 5). These pathways converge downstream, as both caspase-8 and caspase-9 activate caspase-3 (4, 6). Activated caspases cleave specific target proteins, such as protein kinases, cytoskeletal proteins, and DNA repair proteins, initiating apoptosis (6, 7). The characteristic markers of apoptosis include cell and nuclear shrinkage, DNA cleavage into nucleosomal fragments, chromatin aggregation, and apoptotic body formation (8). Because the precisely regulated events of apoptotic cell death are frequently altered in cancers, proteins in the apoptotic pathway can represent good targets for anti-cancer drugs (9).

Inhibitor of apoptosis (IAP)[3] family proteins are negative regulators of apoptosis and were characterized originally as physical inhibitors of caspases (10). Eight IAP proteins have been identified, and each of these proteins contains an ~70-amino acid BIR (baculovirus IAP repeat) domain that mediates protein recognition and interactions (11). Some IAPs include additional domains, such as a RING (really interesting new gene) domain, UBA (ubiquitin-associated domain), or CARD (caspase recruitment domain). Anti-cancer drugs that target IAPs by mimicking the N-terminal IAP binding motif (Ala-Val-Pro-Ile) of Smac are currently in clinical trials for some solid tumors and lymphomas (12).

API5 (apoptosis inhibitor 5), also known as AAC-11 protein (anti-apoptosis clone 11), FIF (fibroblast growth factor 2-interacting factor), and MIG8 (cell migration-inducing gene 8), is a relatively poorly studied apoptosis-inhibiting nuclear protein that does not contain a baculovirus IAP repeat domain. The *API5* gene has been found in animals, protists, and plants (13). The expression of API5 prevents apoptosis following growth factor deprivation, and this protein is up-regulated in various

* This work was supported by the Basic Science Research Program through the National Research Foundation of Korea funded by Ministry of Education, Science, and Technology Grant 2010-0020993 (to B. I. L.).
The atomic coordinates and structure factors (codes 3U0R and 3V6A) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (http://www.rcsb.org/).
[S] This article contains supplemental "Materials and Methods" and Figs. S1–S3.
[1] These authors contributed equally to this work.
[2] To whom correspondence should be addressed: Biomolecular Function Research Branch, Div. of Convergence Technology, Research Inst., National Cancer Ctr., Goyang, Gyeonggi 410-769, Korea. Tel.: 82-31-920-2223; Fax: 82-31-920-2006; E-mail: bilee@ncc.re.kr.

[3] The abbreviations used are: IAP, inhibitor of apoptosis; r.m.s.d., root mean square deviation; PDB, Protein Data Bank; EJC, exon junction complex.

cancer cells (14–19). Non-small cell lung cancer patients whose cancers express API5 have a poorer prognosis than patients with non-API5-expressing cancers, and overexpression of API5 promotes the invasion and inhibits apoptosis in cervical cancer cells (15, 17, 20). It has been suggested that API5 suppresses E2F1 transcription factor-induced apoptosis (21). A recent study also suggested that API5 is phosphorylated by PIM2 kinase and inhibits apoptosis in hepatocellular carcinoma cells through NF-κB (nuclear factor-κB) (22). Various API5 interaction partners have been identified. API5 interacts with high molecular mass forms of fibroblast growth factor 2 (FGF-2), which are involved in cell proliferation and tumorigenesis (16). API5 also binds to and regulates Acinus, a protein involved in chromatin condensation and DNA fragmentation during apoptosis (23). Moreover, the inhibition of API5 increases anticancer drug sensitivity in various cancer cells (23). Recently, the chromatin remodeling enzyme ALC1 (amplified in liver cancer 1) and two DEAD box RNA helicases were identified as binding partners of API5 (13, 24). From these data, API5 has been regarded as a putative metastatic oncogene and therefore represents a therapeutic target for cancer treatment (23). The rational design of inhibitors of API5 will be aided by determination of its three-dimensional structure. Additionally, the lack of sequence similarity between API5 and other known proteins has made it difficult to predict its molecular function. Therefore, structural information about API5 will be helpful in elucidating its molecular function.

We determined the crystal structure of human API5 to elucidate the molecular basis of its anti-apoptotic function. API5 contains a multiple helices, forming HEAT and ARM (Armadillo)-like repeats, which are known to function in protein-protein interactions. This study thus reveals that API5 acts as a mediator of protein interactions.

EXPERIMENTAL PROCEDURES

Protein Expression and Purification—The human *API5* gene (full-length (1–504), API5ΔC (1–454)) was amplified by PCR and cloned into pET28b(+) (Novagen) using NdeI and NotI restriction sites. This construction adds a 21-residue tag including His_6 to the N-terminal of the recombinant protein, facilitating protein purification. The mutants (K251Q, K251R, and K251A) of the *API5* gene were prepared by QuikChange site-directed mutagenesis method using the wild type plasmid as the PCR template. The wild type and mutant proteins were overexpressed in *Escherichia coli* Rosetta2(DE3) cells (Novagen). The cells were grown at 37 °C in 4 liters of Terrific Broth medium to an A_{600} of 0.7, and expression of the recombinant protein was induced with 0.5 mM isopropyl β-D-thiogalactopyranoside at 37 °C. The cells were grown at 37 °C for 16 h after isopropyl β-D-thiogalactopyranoside induction and were harvested by centrifugation at 3000 × *g* for 10 min at 4 °C. The cell pellet was suspended in ice-cold lysis buffer (25 mM Tris-HCl, pH 7.4, 138 mM NaCl, 2 mM KCl, 10% (v/v) glycerol, 1 mM phenylmethylsulfonyl fluoride, and 0.8 μM lysozyme) and homogenized by sonication. The first purification step utilized a Ni^{2+} nitrilotriacetic acid column (Qiagen) for affinity purification via the N-terminal His_6 tag. The eluent was pooled and concentrated. The protein sample was diluted 10-fold with buffer (50 mM Tris-HCl, pH 8.0, 80 mM NaCl, and 1 mM DTT). Further purification was conducted using a HiTrap Q ion exchange chromatography column (GE Healthcare) equilibrated with buffer (50 mM Tris-HCl, pH 8.0, 80 mM NaCl, and 1 mM DTT). The protein was eluted using a linear gradient of 0–1.0 M sodium chloride in the same buffer. The final purification step was gel filtration on a HiLoad 16/60 Superdex 200 prep-grade column (GE Healthcare) equilibrated with buffer containing 20 mM sodium citrate, pH 5.5, 200 mM NaCl, and 1 mM DTT. The purified API5 protein was concentrated to ~20 mg/ml using an YM10 membrane (Millipore). Human α-thrombin (Enzyme Research Laboratories) or porcine trypsin (Promega) was used to remove the fusion tag or to achieve limited proteolysis, respectively.

Crystallization, X-ray Data Collection, and Structure Determination—The best crystals were obtained with a reservoir solution of 10% PEG 6000 and 100 mM bicine, pH 9.0, using the API5ΔC protein. For x-ray diffraction data collection, crystals were transferred to a cryoprotectant solution (10% PEG 6000, 100 mM Bis-Tris, pH 6.0, and 30% ethylene glycol) and mounted on nylon loops. Data of API5ΔCs (wild type and K251Q mutant) were collected using an ADSC Quantum CCD detector at the 4A and 6C experimental stations, Pohang Light Source, Korea and an ADSC Quantum 4 CCD detector at the BL-1A experimental station, Photon Factory, Japan, respectively. Intensity data were processed and scaled using the program HKL2000 (25). Selenium sites were located with SOLVE (26) using two MAD data sets. Initial phases were improved using the program RESOLVE (26). Manual model building was performed using the program COOT (27), and the model was refined with the program PHENIX (28) and Refmac (29), including bulk solvent correction. As the test data for the calculation of R_{free}, 5% of the data were randomly set aside (30). The refined models were evaluated using MolProbity (31).

Analytical Ultracentrifugation—Equilibrium sedimentation experiments were performed using a Beckman ProteomeLab XL-A analytical ultracentrifuge in 20 mM sodium citrate buffer, pH 5.5, containing 200 mM NaCl, at 20 °C. Absorbances of API5 samples were measured at 235, 240, and 280 nm using a two-sector cell at two different speeds (12,000 and 16,000 rpm) and two different concentrations (4.90 and 9.80 μM) with a loading volume of 180 μl. The calculated partial specific volumes at 20 °C were 0.7452 and 0.7416 cm^3/g for API5ΔC and full-length API5, respectively. The buffer density was 1.01424 g/cm^3. For mathematical modeling using non-linear least-squares curve fitting, the fitting function for homogeneous models was used, as described elsewhere (32). The model was selected by examining the weighted sum or square values and weighted root mean square deviation (r.m.s.d.) values.

Cell Viability Test—Jurkat cells were grown in RPMI 1640 media supplemented with 10% fetal bovine serum (WelGENE). Transient transfection was performed with Lipofectamine LTX and Plus Reagent (Invitrogen) according to the manufacturer's instructions. Briefly, 2.3×10^5 Jurkat cells were seeded in the medium onto a 24-well plate. After 1 h, the cells in each well were transfected with 0.6 μg of each DNA construct (full-length API5 wild type, K251R, K251Q, and K251A, respectively). After 3 h, the serum-free medium was exchanged with

RPMI1640 medium containing 10% fetal bovine serum to stabilize the transfected cells. After an additional 12-h incubation, the culture medium was exchanged with serum-free RPMI 1640 to perform a starvation test. Following a 48-h incubation, the cell viability in each well was determined using a Presto cell proliferation assay kit (Invitrogen) according to the manufacturer's instructions.

Immunoprecipitation and Western Blotting—To monitor the oligomerization state of API5 in cells, full-length HA-API5 and 3×FLAG-API5 were co-transfected into HeLa cells and immunoprecipitated. Whole cell lysates prepared in lysis buffer (50 mM Tris-HCl, pH 7.5, 150 mM NaCl, 1% Triton X-100, 0.5% sodium deoxycholate, and 0.1% sodium dodecyl sulfate), were used for immunoprecipitation. Each lysate was mixed with a binding buffer (50 mM Tris-HCl, pH 7.5, 150 mM NaCl, and 1 mM EDTA) supplemented with protease inhibitor cocktails (Roche Applied Science). The mixture was incubated overnight at 4 °C with antibodies (diluted 1:100). A/G-agarose beads were then added, incubated at 4 °C for 3 h, and washed four times with the binding buffer. The immune complex was released from the beads by boiling in sample buffer and was visualized by Western blotting. Western blotting was carried out using established protocols. The primary antibodies used were anti-HA (Santa Cruz Biotechnology), anti-FLAG (Cell Signaling Technologies), and anti-β-actin (Abcam).

To test protein stability in cells, full-length wild type API5 or mutants API5 (K251R and K251Q) was transfected into HeLa cells. At 22 h after transfection, the cells were treated with 10 μg/ml of cycloheximide for 4 h. Then, the cells were collected, and protein levels were analyzed by Western blotting. For the deacetylase inhibition experiments, suberoylanilide hydroxamic acid was pretreated for 20 h to the wild type API5-transfected HeLa cells before cycloheximide treatment.

RESULTS

Structure Determination and Overall Structure—We obtained crystals of full-length and trypsin-treated API5. However, only the crystals of trypsin-treated API5 were suitable for structure determination. From N-terminal amino acid sequencing and MALDI-TOF analysis, we found that the N-terminal fusion tag and ~50 amino acid residues at the C-terminal of recombinant API5 were removed by trypsin and that the trypsin-treated form of API5 corresponded to residues 1–454. The crystal structure of API5ΔC was solved by multi-wavelength anomalous dispersion and refined to 2.50 and 2.60 Å resolution for wild type and K251Q (acetylation mimic) mutant structures, respectively. The refined models of wild type and K251Q mutant exhibited working and free *R*-values of 21.6 and 24.2%, and 21.3 and 25.0%, respectively, with good stereochemistry (Table 1). One subunit of API5ΔC was found to be present in the asymmetric unit of the crystal. The structures of wild type and K251Q mutant are highly similar to each other, with an r.m.s.d. of 0.32 Å for 424 C_α atom pairs (supplemental Fig. S1).

The API5ΔC monomer presented an elongated, all α-helical repeat structure, forming a right-handed superhelix with approximate dimensions of 100 × 35 × 50 Å (Fig. 1*A*). It consisted of 19 α-helices (α1–α19) and two 3_{10} helices. Each helix was observed to be paired with its neighboring helix to form an antiparallel helix pair. Residues 277–278, 322–335, 365–366, 430–431, and 447–454, as well as the four extra N-terminal residues (Gly-Ser-His-Met), artificially added by cloning into pET-28b, were disordered in the crystal. The C-terminal region of API5 (residues 455–504), which contains a nuclear localization signal (Fig. 2) between residues 454–475 and is predicted to be disordered by IUPRED (33), was readily removed by trypsin. Additionally, an electron density map calculated using low resolution (~6 Å) x-ray diffraction data from full-length API5 did not show this C-terminal region, supporting its hypothesized flexibility (data not shown). Several positively charged patches were found on the concave surface of the API5ΔC protein, whereas mainly negatively charged residues were positioned on the convex side (Fig. 1*B*). The positively or negatively charged patches were well conserved in various species (Fig. 1*C*). Basic patches were also found near the putative leucine zipper region (heptad leucine repeat). However, a basic DNA binding region typically followed by a leucine zipper of other leucine repeat proteins, was not found to be present in API5 (Fig. 2).

Searches using the DALI server were carried out to identify structurally similar proteins (34). Because the structure of API5ΔC can be divided into two regions, DALI searches were performed separately for the N-terminal half (α1-α11) and the C-terminal half (α12–α19). The N-terminal half of API5 structurally aligns with the HEAT repeat regions of importin β (PDB code 1O6P, Z-score = 13.1; r.m.s.d., 4.7 Å for 190 structurally aligned residues), the TOG2 domain of Msps (PDB code 2QK2, Z-score = 12.2; r.m.s.d. = 4.5 Å for 176 structurally aligned residues), and protein phosphatase 2A (PDB code 2C5W, Z-score = 11.8, r.m.s.d. = 1.8 Å for 87 structurally aligned residues). Although the sequence identities between the structurally aligned regions and API5 were only 10–12%, the overall folds were quite similar. The C-terminal half of API5 is structurally similar to the core region of the U-box-containing ubiquitin ligase E4 protein Ufd2p (PDB code 2QIZ, residues 461–747). The Z-score and r.m.s.d. values for the aligned 195 amino acid pairs were 8.3 and 4.4 Å, respectively. The amino acid sequences were also quite different between API5 and Ufd2p (lower than 11% sequence identity), although the three-dimensional structures were quite similar, including the characteristic long helix pair (α18-α19). The second most similar protein to the C-terminal half of API5 was p120 catenin (PDB code 3L6Y, Z-score = 7.7, r.m.s.d. = 3.3 Å for the aligned 152 amino acid pairs). Ufd2p and p120 catenin have been classified as ARM-like repeat proteins. These results indicate that the N-terminal half of API5 adopts the characteristic HEAT repeat structure, which is composed of pairs of antiparallel helices, whereas the C-terminal half of API5 shows an ARM-like repeat structure. Superposition of API5 and the structurally similar protein is shown in Fig. 3. The interhelix turns between antiparallel helices are relatively short (between two and four residues) in the HEAT repeats of API5, resulting in a tight fold of the repeat structure, whereas the turns in the ARM-like repeat regions are longer (6 to 22 residues) (Fig. 2).

The amino acid sequence of API5 contains an L*xx*LL motif (16). The L*xx*LL motif forms a short amphipathic α-helix and is

TABLE 1
Data collection and refinement statistics

	Selenomethionine 1		Selenomethionine 2		Wild type	K251Q
Data collection						
X-ray source	PLS-4A		PLS-6C		PLS-6C	PF-BL1A
Space group	$P2_12_12_1$ (Native cell parameters, $a = 46.42$, $b = 88.61$, and $c = 136.29$ Å; $\alpha = \beta = \gamma = 90°$)					
Data set	Selenium (peak)	Selenium (edge)	Selenium (peak)	Selenium (edge)	Native	Native
Resolution (Å)[a]	50.0-3.10 (3.21-3.10)	50.0-3.10 (3.21-3.10)	50.0-3.60 (3.73-3.60)	50.0-3.50 (3.63-3.50)	50.0-2.50 (2.59-2.50)	50.0-2.60 (2.64-2.60)
Redundancy[a]	7.9 (8.1)	7.9 (8.1)	10.3 (8.6)	7.8 (7.0)	4.0 (3.0)	6.7 (6.3)
Completeness (%)[a]	99.4 (100.0)	99.5 (100.0)	99.9 (99.3)	99.8 (98.9)	94.1 (87.4)	99.3 (99.1)
$\langle I \rangle / \langle \sigma_I \rangle$[a]	50.7 (6.7)	49.5 (5.9)	32 (4.7)	31.8 (3.9)	21.8 (2.2)	39.3 (3.8)
R_{merge} (%)[a]	6.5 (34.9)	5.8 (39.4)	12.8 (49.7)	10.4 (51.0)	7.5 (42.1)	7.7 (60.8)
Phasing						
No. of Se sites	5		6			
Figure of merit (before/after RESOLVE)	0.61/0.76					
Refinement						
Resolution (Å)					50.0-2.50	50.0-2.60
No. of reflections					18,082	17,022
R_{work}/R_{free} (%)					21.6/24.2	21.3/25.0
No. of atoms (protein/water)					3,389/119	3,382/60
Average B-factors (protein/water) (Å^2)					47.7/46.4	59.2/49.9
r.m.s.d. from ideal geometry (bond lengths/bond angles)					0.010 Å/1.202°	0.011 Å/1.337°
r.m.s. Z-scores (Bond lengths/bond angles)[b]					0.488/0.544	0.546/0.610
Ramachandran plot (%)[c]						
Favored					95.18	95.17
Outliers					0.00	0.00
Rotamer outliers (%)[c]					0.53	0.79
PDB code					3U0R	3V6A

[a] Values in the parentheses refer to the highest resolution shells.
[b] Values obtained using Refmac (29).
[c] Values obtained using MolProbity (31).

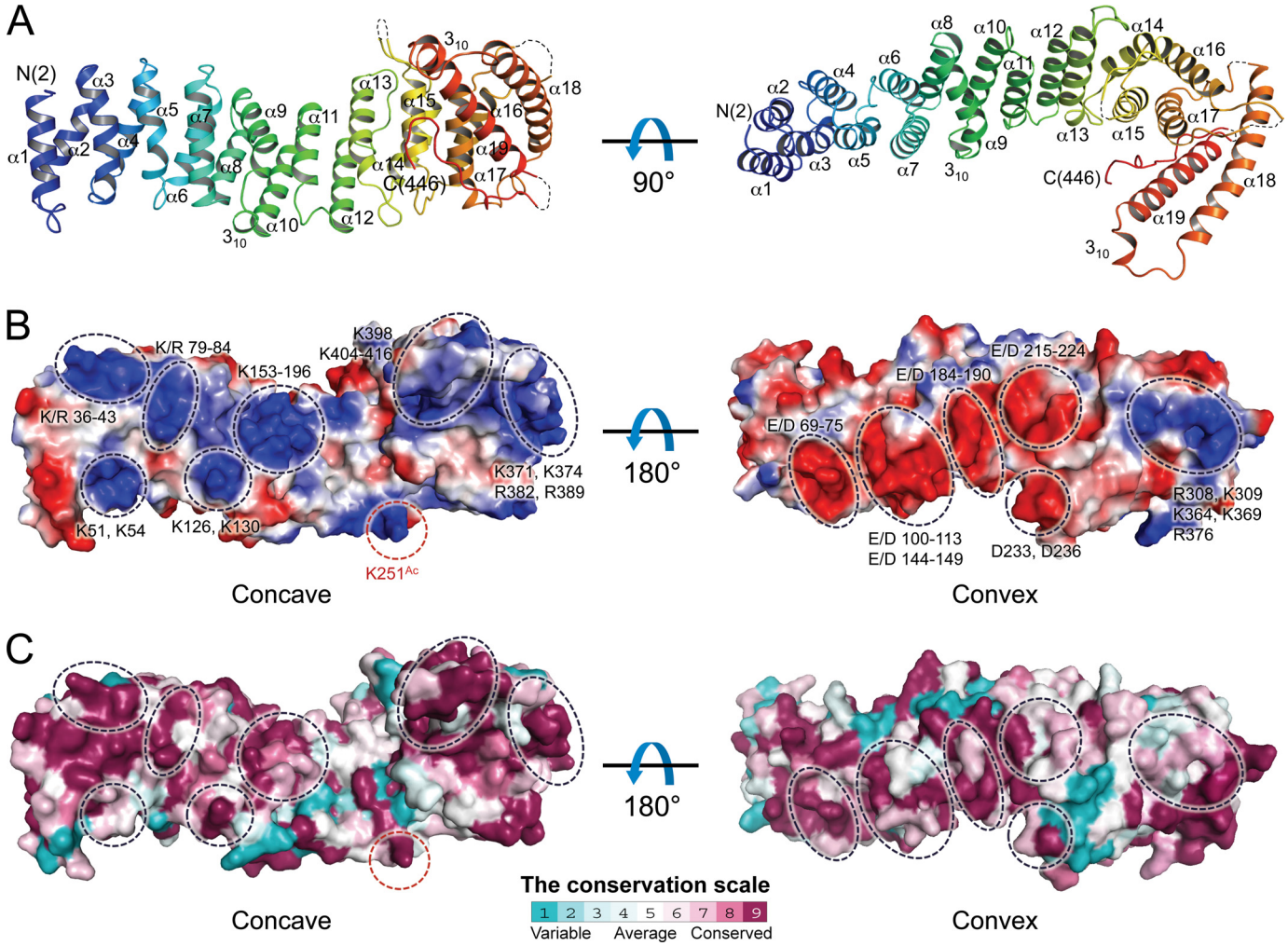


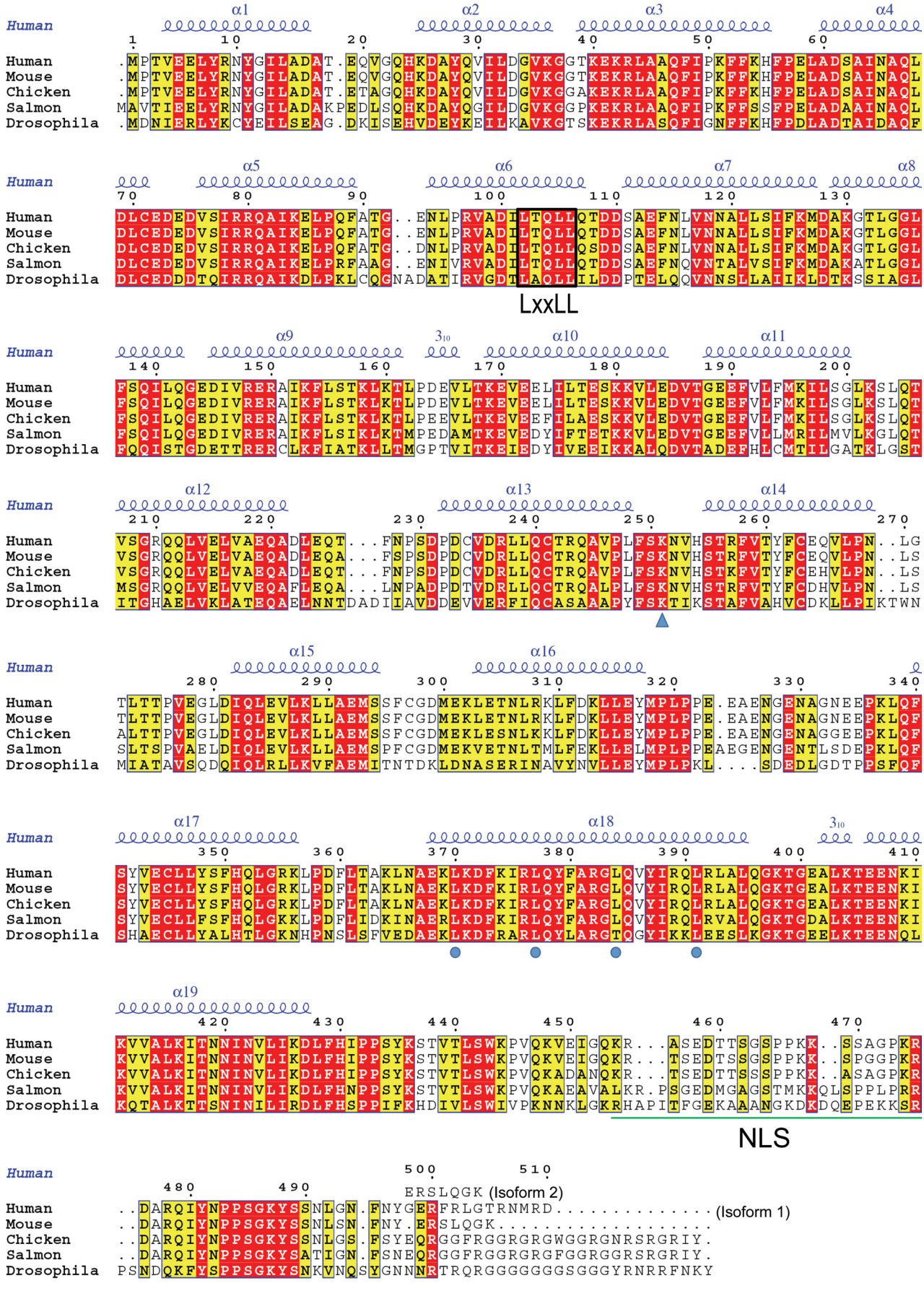
FIGURE 1. **The overall structural representation of API5.** *A*, the ribbon diagram of API5ΔC reveals its elongated, all α-helical structure. *B*, surface charge properties of API5ΔC. The concave region of API5 is highly positively charged, whereas the convex region is negatively charged. The acidic and basic residues in the surface are denoted. Lys-251, shown using *red dashed circles*, is solvent-exposed, allowing access by acetyltransferases or deacetylases. *C*, surface representation colored by amino acid sequence conservation. Surface conservation was generated using ConSurf (57).

usually found in nuclear receptor-cofactor interaction regions (35, 36). The L*xx*LL motif in API5 is positioned in α6 (Figs. 1*A*, 2, 4*A*). However, the three leucine residues in this motif are not located on the surface of the protein and likely do not play any role in protein-protein interactions. Instead, because they are located in the interior of the protein, they may contribute to the stability of the HEAT repeat by forming hydrophobic interactions with neighboring α-helices. Leu-102 interacts with Ala-120, Ile-124 of α7, and Leu-86 of α5. Leu-105 and Leu-106 interact with Val-117 and Phe-114 of α7, respectively (Fig. 4*A*). This finding is consistent with a previous report indicating that the replacement of these conserved leucine residues does not abolish FGF-2 binding (16).

Assessment of API5 Oligomerization and Heptad Leucine Repeat Region—The heptad repeat of leucine residues between residues 370 and 391 in API5 has been predicted to be a leucine zipper without a basic DNA-binding region (14). Therefore, it has been suggested that API5 may form a dimer in solution. However, the crystal structure of API5ΔC indicates that API5 is monomeric and that the putative leucine zipper (α18) does not interact with the corresponding helix of another subunit. Instead, it interacts with α19 of the same subunit (Figs. 1*A* and 4*B*). Two sets of low resolution data were collected from crystals of full-length API5 (~6 Å resolution) or API5ΔC (~4 Å resolution) grown under different crystallization conditions. Protein molecules positioned by the molecular replacement method indicate that both API5 and API5ΔC are monomeric under different crystallization conditions (data not shown).

Because the oligomeric state of API5 in the crystal was found to be different from the predicted dimeric state, the relevant oligomeric state of API5 in solution was further investigated by analytical ultracentrifugation (Fig. 5*A*). The estimated molecular mass (54,949 Da (±234)) agreed with the calculated molecular mass for an API5ΔC monomer, including the fusion tag (53,521 Da). The full-length API5 protein was also found to be monomeric, indicating minimal effects of the C-terminal truncation on the oligomeric structure of API5. The estimated molecular mass of full-length API5 (60,129 Da (±1,164)) agreed with the calculated molecular mass for the full-length API5 monomer including the fusion tag (58,934 Da).

Additionally, we transiently expressed HA-tagged and 3×FLAG-tagged full-length API5 in HeLa cells and performed


α1
α2
α3
α4
Human
Human
Mouse
Chicken
Salmon
Drosophila
α5
α6
α7
α8
LxxLL
α9
3₁₀
α10
α11
α12
α13
α14
α15
α16
α17
α18
3₁₀
α19
NLS
ERSLQGK (Isoform 2)
(Isoform 1)

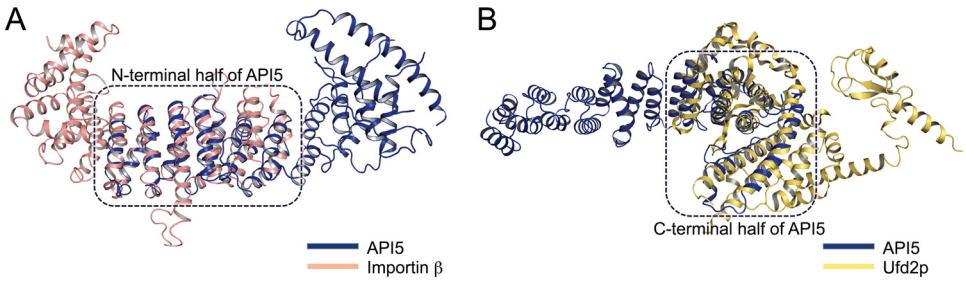


FIGURE 3. **The structural similarity between API5 and HEAT and ARM motif-containing proteins.** *A*, the N-terminal half of API5 is structurally similar to importin β (salmon). *B*, the C-terminal-half of API5 is structurally similar to the core domain of Ufd2p (*yellow*). Superimposed regions are indicated by *dashed boxes*.

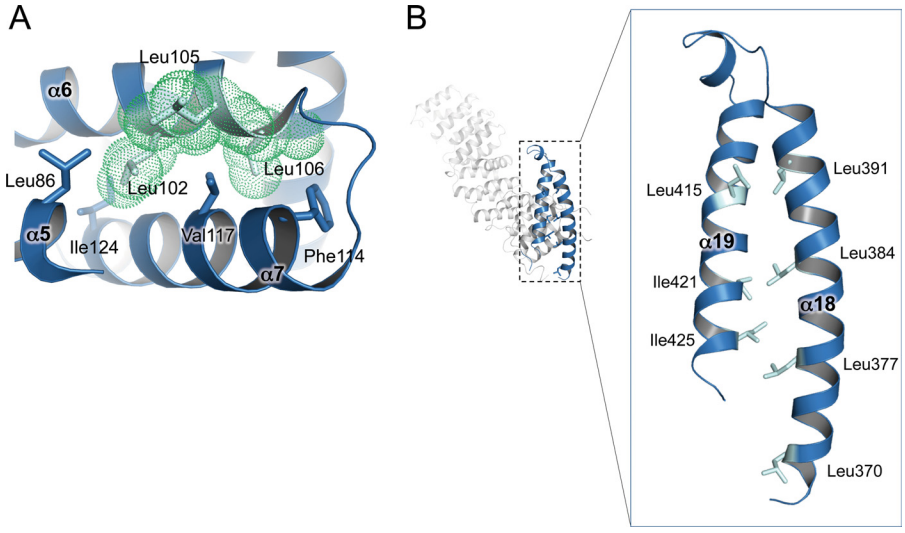


FIGURE 4. **Detailed motif/domain structure of API5.** *A*, L*xx*LL motif and interacting residues. The leucine residues interact with hydrophobic residues on the adjacent helices and are involved in stabilization of the repeated structure. *B*, heptad leucine repeat region. α18 is covered by α19 from the same subunit and does not form a typical leucine zipper. The leucine residues of the heptad repeat and hydrophobic residues on α19 are close to each other (shown in a stick representation).

immunoprecipitation experiments (Fig. 5*B*). No interaction was found between HA-tagged API5 and 3×FLAG-tagged API5. This result shows that API5 does not homodimerize in cells. However, we cannot rule out the possibility that API5 heterodimerize with other leucine zipper-containing proteins under specific cellular conditions. Taken together, we unexpectedly discovered that API5 is monomeric and that its putative leucine zipper helix (α18) does not participate in dimerization. Three of the four leucine residues in the heptad repeat (Leu-377, Leu-384, and Leu-391) exhibit hydrophobic interactions with hydrophobic residues in α19; Leu-377 and Leu-384 interact with Ile-425 and Ile-421, respectively, and Leu-391 is close to Leu-415 (Fig. 4*B*). Leu-370 does not interact with other residues.

Lysine Acetylation and Its Effect on Cancer Cells—A global mass analysis showed that Lys-251 of API5 is acetylated in cells (37). Amino acid sequence alignment showed that this lysine residue is strictly conserved across a number of species (Fig. 2). Lys-251 is positioned in a loop between α13 and α14, which is extended from the protein (Fig. 1*B*). This finding suggests that acetyltransferases or deacetylases can easily access this residue.

API5 plays important roles in the inhibition of cell death under serum deprivation conditions (14, 15). Therefore, we tested whether lysine acetylation affected cell viability 48 h after the transient transfection under serum deprivation conditions (Fig. 6*A*). We made an acetylation-deficient, a constitutive acetylation mimic, and an uncharged mutant (K251R, K251Q, and K251A, respectively) (37, 38). Wild type API5 inhibited cell death by serum starvation compared with untreated or control vector-transfected cells. The K251R mutant exhibited even higher cell viability. However, the K251Q mutant did not inhibit apoptosis efficiently compared with the wild type API5 and the K251R mutant. The K251A mutant did not have anti-apoptotic function. This result means that the inhibition of apoptosis by API5 can be negatively regulated by Lys-251 acetylation. We tried to find the structural basis for anti-apoptotic function by lysine acetylation from the structure of K251Q mutant. However, no significant structural differences were found (supplemental Fig. S1).

FIGURE 2. **Amino acid sequence alignment of API5 homologs from different species.** The L*xx*LL motif is marked with a *black box*. The conserved leucine residues of the heptad leucine repeat are marked by *blue circles*, and the acetylation site Lys-251 is marked by a *blue triangle*. The L*xx*LL motif and nuclear localization signal (*NLS*) sequence are marked. The amino acid sequence for isoform 2 of human API5, which was used in this study, is also denoted above the sequence. The figure was prepared using ESPript.

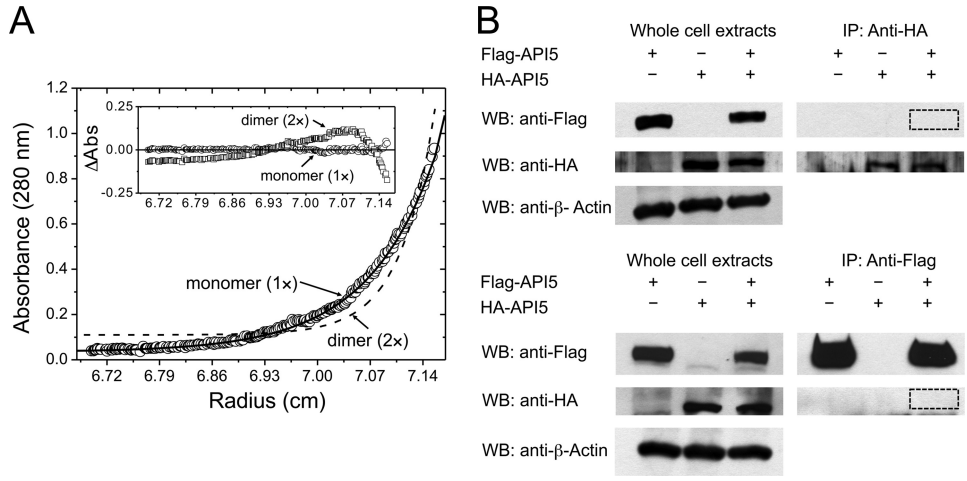


FIGURE 5. **Oligomerization of API5.** *A*, analytical ultracentrifugation analysis. Distributions of the residuals according to both a monomer (1×, *circle*) and a dimer (2×, *square*) models. The random distributions of residuals for the monomer (1×) model indicate that API5ΔC exists as a homogeneous monomer in solution. *B*, immunoprecipitation (*IP*) and Western blot (*WB*) analysis with anti-FLAG and anti-HA API5. No interaction was found between FLAG-API5 and HA-API5, ruling out the possibility of API5 homodimerization.

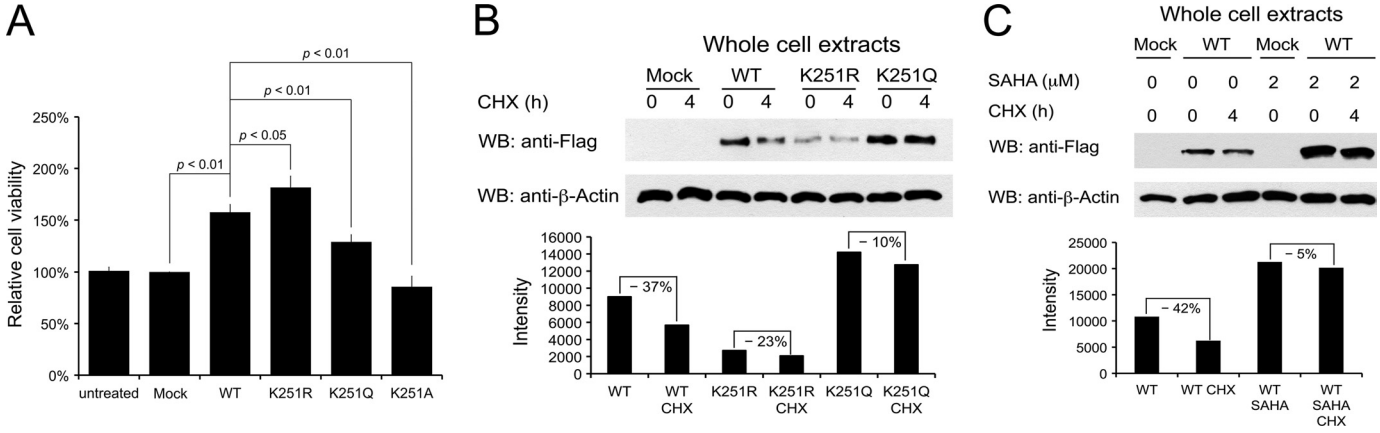


FIGURE 6. **Effects of API5 acetylation at Lys-251.** Wild type API5, an acetylation-deficient mutant (K251R), a constitutive acetylation mimic mutant (K251Q), and an uncharged mutant (K251A) were transiently expressed. *A*, effects on cell viability under conditions of serum deprivation. The K251Q and K251A mutants do not inhibit apoptosis efficiently. *B*, effects of Lys-251 acetylation on protein stability. Cycloheximide (*CHX*) treatment was used to inhibit protein synthesis, and Western blot (*WB*)/normalized densitiometry analysis was performed to monitor protein stability in cells. *C*, effects on protein stability when deacetylase inhibitor, suberoylanilide hydroxamic acid (*SAHA*), was co-treated with cycloheximide to wild type API5-transfected cells.

We also tested whether lysine acetylation affected API5 protein stability. After 22 h of transfection, protein synthesis was stopped by treatment with cycloheximide, and the protein level was monitored by Western blot analysis after 4 h of treatment. The K251Q mutant was more stable than the wild type API5 and the K251R mutant (Fig. 6*B*). Similar results were also found when deacetylase inhibitor suberoylanilide hydroxamic acid was treated (Fig. 6*C*). Therefore, we conclude that the regulation of API5 function by lysine acetylation may occur via effects on protein stability. However, we could not detect any ubiquitination of API5 at Lys-251, suggesting the proteasome-independent degradation pathway (data not shown).

The cellular localization of API5 also was monitored to determine whether lysine acetylation affected the nuclear transport of API5. All of the wild type and mutants API5 (K251R and K251Q) were observed in the cell nuclei, implying that lysine acetylation did not affect cellular localization (supplemental Fig. S2). The analysis of the API5 and FGF-2 interaction detected by *in vitro* immunoprecipitation method using the purified recombinant API5 and FGF-2 did not show any significant FGF-2 binding differences between wild type and mutant API5s (supplemental Fig. S3).

DISCUSSION

In this study, we determined the crystal structure of human API5 and found that API5 contains protein-protein interaction modules, such as HEAT and ARM repeats (Fig. 1*A*). Different types of helix repeat, including HEAT, ARM, ankyrin, tetratrico peptide, and leucine-rich variant repeats, have been found in many proteins and have diverse functional roles in mediating protein-protein interactions (39). The elongated structure of API5 is well suited for interactions with multiple binding partners, similar to the roles of other repeat proteins (40).

The structures of various HEAT and ARM repeat-containing proteins in complexes with other proteins have been elucidated. Ran-importin β, protein phosphatase 2 holoenzyme, and the Cand1-Cul1-Roc1 complex are good examples of HEAT repeat complex structures (importin β, the A subunit of protein phosphatase 2 holoenzyme, and Cand1 are HEAT repeat-containing proteins) (41–43). Several crystal structures of the

ARM repeat-containing protein (*e.g.* β-catenin) in complex with the interacting proteins have been determined (44–47). In many curved HEAT repeat proteins, protein-protein interactions are mediated by the concave surface of the protein, regardless of the surface charge distribution, and the protein-protein interaction surface involving the ARM repeat proteins usually span the entire range of the repeat. The charged concave or convex side of the API5 is well conserved across a number of species. This suggests that binding partners can bind to these regions.

It has been reported that high molecular mass FGF-2 interacts with API5 in the nucleus, and immunoprecipitation results suggest that two separate regions corresponding to three helices (α6 and α15–α16) of API5 are important for FGF-2 binding (16). The α6 is exposed to the convex surface of API5, and α15-α16 are exposed on both the concave and convex surfaces. Analyses of surface properties indicated that the convex side of API5 is highly negatively charged, whereas FGF-2 is highly positively charged (Fig. 1*B*). Although hydrophobic interactions may also be important for binding, we predict that the positively charged high molecular mass FGF-2 likely binds to the negatively charged convex surface of API5, as determined from previous immunoprecipitation results and surface electrostatic potential (16).

It has been reported that API5 binds to Acinus and protects it from cleavage by caspase-3, resulting in an inhibition of apoptosis (23). Acinus has been identified as a component of an ASAP (apoptosis and splicing-associated protein) complex and an exon junction complex (EJC) (48, 49). API5 and Acinus interact with each other through the heptad leucine repeat region of API5, which seems to be important for API5 function and may therefore represent a potential therapeutic target for anti-cancer drugs (23). Mutation of leucine residues (Leu-384 and Leu-391) in the heptad leucine repeat region abrogates the anti-apoptotic effects of API5 (14, 23). Because Leu-384 and Leu-391 contribute to hydrophobic interactions with other hydrophobic residues in α19, mutation of these residues might cause a distortion of the local structure of API5, thus affecting protein-protein interactions, rather than inhibiting the dimerization. Because API5 is monomeric and the heptad leucine repeat of API5 is structurally distinct from that of other leucine zippers, API5 provides a unique opportunity for anti-cancer drug discovery through targeting of this region. However, further studies and target validation will be required to develop API5-targeted inhibitors.

It is known that Nϵ-acetylation at lysine residues affects DNA binding, protein-protein interaction, cellular localization, ubiquitination, phosphorylation, and protein stability (50). A global analysis of lysine acetylation using high resolution mass spectrometry showed that Lys-251 of API5 is acetylated (37). Our data clearly showed that lysine acetylation impacts cell survival under serum deprivation conditions, suggesting an inhibition of function. Therefore, we conclude that API5 function is negatively regulated by acetylation at Lys-251. Inhibition of function via acetylation has also been found in HSP90 and 14-3-3 proteins (37, 38). We determined that lysine acetylation affects API5 protein stability, which has been found for p53 as well (51). However, we could not detect ubiquitination of API5, and a detailed mechanism for degradation of API5 has not yet been established. We predict that API5 may be readily acetylated by an acetyltransferase(s) after protein synthesis and that API5 exists in an inactive and stable state in cells until certain deacetylases can activate API5 through deacetylation. However, activated API5 appears to be unstable and easily degraded in cells. Although the acetyltrasferases or deacetylases that act on API5 have not yet been identified, these enzymes are likely to be important regulatory factors for the function of API5. Additionally, we predict that certain histone deacetylase inhibitors may also be effective in regulating API5 function. We could not found any significant differences in API5-FGF-2 interaction and cellular localization from the lysine acetylation mimic/deficient mutants. However, it is still possible that lysine acetylation of API5 will affect protein-protein interactions between API5 and other API5 interacting partners to induce anti-apoptosis function. We can easily predict that lysine acetylation can inhibit the binding of API5 to its binding partners for its proper function. However, it is also quite likely that lysine acetylated API5 can be recognized by new binding partners, such as bromodomain containing proteins, resulting in a negative regulation of apoptosis (52). Hence, further investigations are needed to elucidate the detailed mechanism for functional regulation by lysine acetylation.

A recent report suggested that the rice API5 homolog interacts with two DEAD-box RNA helicases AIP1/2 (API5-interacting protein 1/2), thereby regulating programmed cell death in the tapetum during the development of male gametophytes (13). The human homolog of rice AIP1/2 is the ATP-dependent RNA helicases UAP56. This protein is involved in pre-mRNA splicing and mRNA nuclear export (53). Interestingly, hUAP56 is a component of large protein complexes, such as the EJC or TREX (Transcription-export) complex (54, 55). Particularly, Acinus, another API5 binding protein, is also a component of EJC complex, implying the functional link between API5 and EJC complex. Another study showed that API5 interacts with the chromatin-remodeling enzyme ALC1 (also known as CHD1L), which is a member of the SNF2 family containing a DEAH box helicase domain (24). ALC1 plays important roles in promoting cell proliferation and inhibits apoptosis in hepatocellular carcinoma (56). ALC1 also has been shown to interact with various DNA repair proteins such as DNA-PKCs, Ku, PARP1 (Poly [ADP-ribose] polymerase 1), XRCC (X-ray repair cross-complementing protein), and APLF (Aprataxin PNK-like factor) as well as API5 (24), implying that API5 may also mediate protein-protein interactions in DNA damage response protein complexes. Taken together, the interactions between API5 and EJC component proteins and chromatin-remodeling enzyme such as ALC1 suggest the possible roles of API5 at a transcriptional level.

In conclusion, the protein structure obtained in this study reveals that API5 possesses protein-protein interaction modules that likely mediate interactions with partners such as high molecular mass FGF-2, Acinus, AIP1/2, and ALC1. These interacting proteins are usually components of multiprotein complexes. Therefore, API5 may serve as a scaffold for multiprotein complexes, and the identification of interacting protein partners will help to illuminate the functional roles of API5.

Furthermore, since we show that the API5 function is regulated by lysine acetylation, identifying the lysine acetyltransferases and deacetylases that act on API5 should also yield a better understanding of the mechanisms that regulate API5 function.

Acknowledgments—We thank the staff at the Pohang Light Source (beamlines 4A and 6C) and the Photon Factory (BL-1A) for assistance with synchrotron data collection.

REFERENCES

1. Meier, P., Finch, A., and Evan, G. (2000) Apoptosis in development. *Nature* **407,** 796–801
2. Danial, N. N., and Korsmeyer, S. J. (2004) Cell death: Critical control points. *Cell* **116,** 205–219
3. Igney, F. H., and Krammer, P. H. (2002) Death and anti-death: Tumor resistance to apoptosis. *Nat. Rev. Cancer* **2,** 277–288
4. Yang, S. Y., Sales, K. M., Fuller, B., Seifalian, A. M., and Winslet, M. C. (2009) Apoptosis and colorectal cancer: Implications for therapy. *Trends Mol. Med.* **15,** 225–233
5. Altieri, D. C. (2010) Survivin and IAP proteins in cell death mechanisms. *Biochem. J.* **430,** 199–205
6. Ghobrial, I. M., Witzig, T. E., and Adjei, A. A. (2005) Targeting apoptosis pathways in cancer therapy. *CA Cancer J. Clin.* **55,** 178–194
7. Alnemri, E. S., Livingston, D. J., Nicholson, D. W., Salvesen, G., Thornberry, N. A., Wong, W. W., and Yuan, J. (1996) Human ICE/CED-3 protease nomenclature. *Cell* **87,** 171
8. Wyllie, A. H., Kerr, J. F., and Currie, A. R. (1980) Cell death: The significance of apoptosis. *Int. Rev. Cytol.* **68,** 251–306
9. Kasibhatla, S., and Tseng, B. (2003) Why target apoptosis in cancer treatment? *Mol. Cancer Ther.* **2,** 573–580
10. Salvesen, G. S., and Duckett, C. S. (2002) IAP proteins: Blocking the road to death's door. *Nat. Rev. Mol. Cell Biol.* **3,** 401–410
11. Srinivasula, S. M., and Ashwell, J. D. (2008) IAPs: What's in a name? *Mol. Cell* **30,** 123–135
12. Gyrd-Hansen, M., and Meier, P. (2010) IAPs: From caspase inhibitors to modulators of NF-κB, inflammation, and cancer. *Nat. Rev. Cancer* **10,** 561–574
13. Li, X., Gao, X., Wei, Y., Deng, L., Ouyang, Y., Chen, G., Li, X., Zhang, Q., and Wu, C. (2011) Rice APOPTOSIS INHIBITOR5 coupled with two DEAD-box adenosine 5′-triphosphate-dependent RNA helicases regulates tapetum degeneration. *Plant Cell* **23,** 1416–1434
14. Tewari, M., Yu, M., Ross, B., Dean, C., Giordano, A., and Rubin, R. (1997) AAC-11, a novel cDNA that inhibits apoptosis after growth factor withdrawal. *Cancer Res.* **57,** 4063–4069
15. Kim, J. W., Cho, H. S., Kim, J. H., Hur, S. Y., Kim, T. E., Lee, J. M., Kim, I. K., and Namkoong, S. E. (2000) AAC-11 overexpression induces invasion and protects cervical cancer cells from apoptosis. *Lab. Invest.* **80,** 587–594
16. Van den Berghe, L., Laurell, H., Huez, I., Zanibellato, C., Prats, H., and Bugler, B. (2000) FIF (fibroblast growth factor-2 (FGF-2)-interacting-factor), a nuclear putatively antiapoptotic factor, interacts specifically with FGF-2. *Mol. Endocrinol.* **14,** 1709–1724
17. Sasaki, H., Moriyama, S., Yukiue, H., Kobayashi, Y., Nakashima, Y., Kaji, M., Fukai, I., Kiriyama, M., Yamakawa, Y., and Fujii, Y. (2001) Expression of the antiapoptosis gene, AAC-11, as a prognosis marker in non-small cell lung cancer. *Lung Cancer* **34,** 53–57
18. Clegg, N., Ferguson, C., True, L. D., Arnold, H., Moorman, A., Quinn, J. E., Vessella, R. L., and Nelson, P. S. (2003) Molecular characterization of prostatic small-cell neuroendocrine carcinoma. *Prostate* **55,** 55–64
19. Krejci, P., Pejchalova, K., Rosenbloom, B. E., Rosenfelt, F. P., Tran, E. L., Laurell, H., and Wilcox, W. R. (2007) The antiapoptotic protein Api5 and its partner, high molecular weight FGF2, are up-regulated in B cell chronic lymphoid leukemia. *J. Leukoc. Biol.* **82,** 1363–1364
20. Wang, Z., Liu, H., Liu, B., Ma, W., Xue, X., Chen, J., and Zhou, Q. (2010) Gene expression levels of CSNK1A1 and AAC-11, but not NME1, in tumor tissues as prognostic factors in NSCLC patients. *Med. Sci. Monit.* **16,** CR357–364
21. Morris, E. J., Michaud, W. A., Ji, J. Y., Moon, N. S., Rocco, J. W., and Dyson, N. J. (2006) Functional identification of Api5 as a suppressor of E2F-dependent apoptosis *in vivo*. *PLoS Genet.* **2,** e196
22. Ren, K., Zhang, W., Shi, Y., and Gong, J. (2010) Pim-2 activates API-5 to inhibit the apoptosis of hepatocellular carcinoma cells through NF-κB pathway. *Pathol. Oncol. Res.* **16,** 229–237
23. Rigou, P., Piddubnyak, V., Faye, A., Rain, J. C., Michel, L., Calvo, F., and Poyet, J. L. (2009) The antiapoptotic protein AAC-11 interacts with and regulates Acinus-mediated DNA fragmentation. *EMBO J.* **28,** 1576–1588
24. Ahel, D., Horejsí, Z., Wiechens, N., Polo, S. E., Garcia-Wilson, E., Ahel, I., Flynn, H., Skehel, M., West, S. C., Jackson, S. P., Owen-Hughes, T., and Boulton, S. J. (2009) Poly(ADP-ribose)-dependent regulation of DNA repair by the chromatin remodeling enzyme ALC1. *Science* **325,** 1240–1243
25. Otwinowski, Z., and Minor, W. (1997) Processing of x-ray diffraction data collected in oscillation mode. *Methods in Enzymology* **276,** 307–326
26. Terwilliger, T. C., and Berendzen, J. (1999) Automated MAD and MIR structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **55,** 849–861
27. Emsley, P., and Cowtan, K. (2004) Coot: Model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60,** 2126–2132
28. Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. (2010) PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66,** 213–221
29. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) Refinement of macromolecular structures by the maximum likelihood method. *Acta Crystallogr. D Biol. Crystallogr.* **53,** 240–255
30. Brünger, A. T. (1992) Free *R* value: A novel statistical quantity for assessing the accuracy of crystal structures. *Nature* **355,** 472–475
31. Chen, V. B., Arendall, W. B., 3rd, Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., and Richardson, D. C. (2010) MolProbity: All-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **66,** 12–21
32. Jung, W. S., Hong, C. K., Lee, S., Kim, C. S., Kim, S. J., Kim, S. I., and Rhee, S. (2007) Structural and functional insights into intramolecular fructosyl transfer by inulin fructotransferase. *J. Biol. Chem.* **282,** 8414–8423
33. Dosztányi, Z., Csizmok, V., Tompa, P., and Simon, I. (2005) IUPred: Web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics* **21,** 3433–3434
34. Holm, L., and Rosenström, P. (2010) Dali server: Conservation mapping in 3D. *Nucleic Acids Res.* **38,** W545–549
35. Heery, D. M., Kalkhoven, E., Hoare, S., and Parker, M. G. (1997) A signature motif in transcriptional co-activators mediates binding to nuclear receptors. *Nature* **387,** 733–736
36. Plevin, M. J., Mills, M. M., and Ikura, M. (2005) The L*xx*LL motif: A multifunctional binding sequence in transcriptional regulation. *Trends Biochem. Sci.* **30,** 66–69
37. Choudhary, C., Kumar, C., Gnad, F., Nielsen, M. L., Rehman, M., Walther, T. C., Olsen, J. V., and Mann, M. (2009) Lysine acetylation targets protein complexes and co-regulates major cellular functions. *Science* **325,** 834–840
38. Scroggins, B. T., Robzyk, K., Wang, D., Marcu, M. G., Tsutsumi, S., Beebe, K., Cotter, R. J., Felts, S., Toft, D., Karnitz, L., Rosen, N., and Neckers, L. (2007) An acetylation site in the middle domain of Hsp90 regulates chaperone function. *Mol. Cell* **25,** 151–159
39. Groves, M. R., and Barford, D. (1999) Topological characteristics of helical repeat proteins. *Curr. Opin. Struct. Biol.* **9,** 383–389
40. Tu, D., Li, W., Ye, Y., and Brunger, A. T. (2007) Structure and function of the yeast U-box-containing ubiquitin ligase Ufd2p. *Proc. Natl. Acad. Sci. U.S.A.* **104,** 15599–15606
41. Vetter, I. R., Arndt, A., Kutay, U., Görlich, D., and Wittinghofer, A. (1999) Structural view of the Ran-Importin β interaction at 2.3 A resolution. *Cell* **97,** 635–646
42. Cho, U. S., and Xu, W. (2007) Crystal structure of a protein phosphatase 2A heterotrimeric holoenzyme. *Nature* **445,** 53–57

43. Goldenberg, S. J., Cascio, T. C., Shumway, S. D., Garbutt, K. C., Liu, J., Xiong, Y., and Zheng, N. (2004) Structure of the Cand1-Cul1-Roc1 complex reveals regulatory mechanisms for the assembly of the multisubunit cullin-dependent ubiquitin ligases. *Cell* **119,** 517–528
44. Choi, H. J., Gross, J. C., Pokutta, S., and Weis, W. I. (2009) Interactions of plakoglobin and β-catenin with desmosomal cadherins: Basis of selective exclusion of α- and β-catenin from desmosomes. *J. Biol. Chem.* **284,** 31776–31788
45. Sampietro, J., Dahlberg, C. L., Cho, U. S., Hinds, T. R., Kimelman, D., and Xu, W. (2006) Crystal structure of a β-catenin-BCL9-Tcf4 complex. *Mol. Cell* **24,** 293–300
46. Xing, Y., Clements, W. K., Kimelman, D., and Xu, W. (2003) Crystal structure of a β-catenin-axin complex suggests a mechanism for the β-catenin destruction complex. *Genes Dev.* **17,** 2753–2764
47. Huber, A. H., and Weis, W. I. (2001) The structure of the β-catenin-E-cadherin complex and the molecular basis of diverse ligand recognition by β-catenin. *Cell* **105,** 391–402
48. Schwerk, C., Prasad, J., Degenhardt, K., Erdjument-Bromage, H., White, E., Tempst, P., Kidd, V. J., Manley, J. L., Lahti, J. M., and Reinberg, D. (2003) ASAP, a novel protein complex involved in RNA processing and apoptosis. *Mol. Cell Biol.* **23,** 2981–2990
49. Tange, T. Ø., Shibuya, T., Jurica, M. S., and Moore, M. J. (2005) Biochemical analysis of the EJC reveals two new factors and a stable tetrameric protein core. *RNA* **11,** 1869–1883
50. Yang, X. J. (2004) Lysine acetylation and the bromodomain: A new partnership for signaling. *Bioessays* **26,** 1076–1087
51. Yang, X. J., and Seto, E. (2008) Lysine acetylation: Codified cross-talk with other post-translational modifications. *Mol. Cell* **31,** 449–461
52. Zeng, L., and Zhou, M. M. (2002) Bromodomain: An acetyllysine binding domain. *FEBS Lett.* **513,** 124–128
53. Shen, H. (2009) UAP56, a key player with surprisingly diverse roles in pre-mRNA splicing and nuclear export. *BMB Rep* **42,** 185–188
54. Strässer, K., Masuda, S., Mason, P., Pfannstiel, J., Oppizzi, M., Rodriguez-Navarro, S., Rondón, A. G., Aguilera, A., Struhl, K., Reed, R., and Hurt, E. (2002) TREX is a conserved complex coupling transcription with messenger RNA export. *Nature* **417,** 304–308
55. Le Hir, H., and Andersen, G. R. (2008) Structural insights into the exon junction complex. *Curr. Opin. Struct. Biol.* **18,** 112–119
56. Ma, N. F., Hu, L., Fung, J. M., Xie, D., Zheng, B. J., Chen, L., Tang, D. J., Fu, L., Wu, Z., Chen, M., Fang, Y., and Guan, X. Y. (2008) Isolation and characterization of a novel oncogene, amplified in liver cancer 1, within a commonly amplified region at 1q21 in hepatocellular carcinoma. *Hepatology* **47,** 503–510
57. Ashkenazy, H., Erez, E., Martz, E., Pupko, T., and Ben-Tal, N. (2010) ConSurf 2010: Calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic Acids Res.* **38,** W529–533